

Centre for Immune Regulation
Institute of Immunology
Faculty Division Rikshospitalet
University of Oslo
Oslo
Norway

**Exploring peptide binding to the disease associated
HLA-DQ2.5 molecule by the use of peptide libraries**

**Doctoral thesis by
Ulrike Jüse**



Oslo 2010

© **Ulrike Jüse, 2010**

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1009*

ISBN 978-82-8072-545-5

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Inger Sandved Anfinssen.
Printed in Norway: AiT e-dit AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Contents

| | |
|--|----|
| Acknowledgements | 4 |
| Abbreviations | 5 |
| List of Papers..... | 6 |
| 1. Introduction | 7 |
| 1.1. MHC class II | 8 |
| 1.1.1. The MHC:peptide complex | 9 |
| 1.1.2. The peptide binding motif of HLA-DQ2.5 | 10 |
| 1.1.3. Gluten derived peptides presented by HLA-DQ2.5..... | 12 |
| 1.1.4. Detergent-solubilized and water soluble MHC class II molecules | 14 |
| 1.2. Celiac disease..... | 14 |
| 1.2.1. Pathogenesis | 15 |
| 1.2.2. Clinical diagnosis | 18 |
| 1.2.3. Genetics | 18 |
| 1.2.4. Treatment possibilities today and in the future..... | 19 |
| 1.2.5. Blocking presentation and/or recognition of gluten derived peptides..... | 21 |
| 2. Aims of this study..... | 22 |
| 3. Summary of the papers | 23 |
| 4. Methodological considerations..... | 25 |
| 4.1. Peptide library approach | 25 |
| 4.2. Synthesis of peptides and peptide libraries | 27 |
| 4.3. HLA-DQ2.5 as full length molecule or soluble construct..... | 28 |
| 4.4. Isolation and identification of high affinity HLA ligands from peptide libraries..... | 29 |
| 4.5. Peptide binding assays | 30 |
| 5. Discussion | 32 |
| 5.1. Interfering with the presentation of immunogenic epitopes by HLA..... | 32 |
| 5.2. Methods for identification of HLA binding motifs and optimal binders | 33 |
| 5.3. Optimizing fitting to increase binding affinity to HLA-DQ2.5..... | 35 |
| 5.3.1. Optimizing anchors..... | 35 |
| 5.3.2. Influence of non-anchor residues..... | 36 |
| 5.4. Advantages of a water soluble approach..... | 37 |
| 5.5. Limitations of the soluble peptide library approach applied..... | 38 |
| 5.5.1. Sliding of the binding frame | 38 |
| 5.5.2. Libraries with too high complexity..... | 39 |
| 5.6. Further optimization of the soluble peptide library approach | 40 |
| 6. Conclusion..... | 41 |
| References | 43 |
| Errata | 53 |

Acknowledgements

The present work was carried out at the Centre for Immune Regulation, Institute of Immunology, Faculty Division Rikshospitalet, University of Oslo, during the period 2006-2009.

This work has been funded as part of a European Commission Marie Curie Research Training Network and by a smaller grant from Helse Sør Øst for which I am most grateful.

This thesis would not have been possible without the constant encouragement and important advices from Ludvig M. Sollid, who has been my main supervisor during this time period. His incredibly fast and always constructive feedback has been very valuable to me.

I would also like to express my gratefulness to Burkhard Fleckenstein, who has been my co-supervisor. Especially his detailed knowledge in peptide chemistry, biochemistry and proteomics were invaluable to me.

I owe my most sincere gratitude to all the gutfeelers, it was you guys who made this tough time fun!

An especially warm thank you goes to the office A2029 members and pellet coffee lovers, Jorunn Stamnæs, Lars-Egil Fallang, Elin Bergseng, Michael Bodd, Roberto DiNiro and Silja S. Amundsen, for numerous valuable discussions and for all the support and backup during the last years.

I am indebted to my colleagues and CD-group members for their help and support: Bjørg Simonsen, Marie K. Johannesen, Stig Tollefsen, Melinda Ráki, Shuo-Wang Qiao, Ann-Christin Røberg Beitnes, Luka Mesin, Rasmus Anders Holm, Anders Fallang, Sylvie Pollmann, Maria Stensland, Siri Dørum, Astrid Tuttøren, Marit Jørgensen, Ingrid Olsen, Øyvind Molberg, Knut E. A. Lundin and Margit Brottveit. I am very proud to be part of this fantastic and inspiring research group.

I would also like to acknowledge the contribution of my co-authors, Magnus Arntzen, Peter Højrup, Frits Koning, and Yvonne van de Wal.

I am very grateful for all support I received from Statens legemiddelverk and my new colleagues during the last year, allowing me take the time I needed to finish this work.

And finally, I like to express my deepest thanks to family, friends and especially Håvard, for unconditional help, patience and love.

Ulrike

Abbreviations

| | |
|---------------|---|
| APC | Antigen presenting cell |
| CD | Celiac disease |
| CLIP | Class II-associated invariant chain peptide |
| EBV | Epstein-barr-virus |
| ER | Endoplasmic reticulum |
| Fmoc | 9-Fluorenyl-methoxycarbonyl |
| GFD | Gluten free diet |
| HLA | Human leukocyte antigen |
| HPLC | High performance liquid chromatography |
| Hsp | Heat shock protein |
| IEL | Intraepithelial lymphocyte |
| IFN- γ | Interferon gamma |
| Ig | Immunoglobulin |
| li | Invariant chain |
| IL | Interleukin |
| MALDI | Matrix assisted laser desorption/ionisation |
| MHC | Major histocompatibility complex |
| MIIC | MHC class II compartment |
| MS | Mass spectrometry |
| nLC | nano liquid chromatography |
| PBS | Phosphate buffered saline |
| PEP | Prolyl endopeptidase |
| sDQ2 | water soluble HLA-DQ2.5 |
| SDS | Sodium dodecylsulfate |
| TCR | T-cell receptor |
| TG2 | Transglutaminase type 2 |
| TOF | Time of flight |

List of Papers

Paper I

Soluble HLA-DQ2 expressed in S2 cells copurifies with a high affinity insect cell derived protein

Ulrike Jüse, Burkhard Fleckenstein, Elin Bergseng, Ludvig M. Sollid

Immunogenetics 2009 Feb;61(2):81-9

Paper II:

Design of new high-affinity peptide ligands for human leukocyte antigen-DQ2 using a positional scanning peptide library

Ulrike Jüse, Yvonne van de Wal, Frits Koning, Ludvig M. Sollid, Burkhard Fleckenstein

Hum Immunol. 2010 May;71(5):475-81

Paper III - manuscript

Assessing high affinity binding to HLA-DQ2.5 by a novel library-based approach

Ulrike Jüse, Magnus Arntzen, Peter Højrup, Burkhard Fleckenstein and Ludvig M. Sollid

Manuscript (submitted to Bioorganic & Medical Chemistry, 03/2010)

1. Introduction

Our immune system protects us from diseases caused by pathogenic organisms, foreign cells and tissues and from the development of tumors. It is a system of highly complex processes, specialized cells and tissues. Disorders in the immune system can cause diseases. The so called immunodeficiency diseases are characterized by a less active or inactive immune system. This is causing the host to be especially vulnerable to infections. Autoimmunity on the other side is characterized by an overactive immune system, recognizing normal tissue and proteins as *non-self*. This is causing the immune system to destruct healthy own cells and tissues. Hypersensitivity diseases are damaging reactions of a normal immune system, which require a prior sensitization of the host.

The healthy immune system of vertebrates acts in a very specific way. Its activity is inducible and can be adapted in order to ensure highest efficiency. In general the network can be divided into two main parts, the innate and the adaptive immunity. The innate immune system is the evolutionary older part and the so called first line defense. It is a fast but relatively unspecific response to invading pathogens. This arm of the immune system does not have a memory and can not adapt and be more efficient upon second-time encounter of the same antigen. The main components are the surface barriers, the complement and immune cells like phagocytes and natural killer cells. These are readily present also before the onset of an infection. The response of the adaptive immune system is dependent on highly specific antigen receptors expressed by the major players, the B and T lymphocytes. Upon activation, the antigen specific cells have to undergo clonal expansion, causing the typical delay in the onset of an adaptive response. Another major difference is the generation of memory, allowing the adaptive immune system to remember pathogens and to react stronger and more efficiently upon a second encounter.

Major histocompatibility complex (MHC) molecules play a central role in the cell mediated immune response. On the surface of vertebrate, nucleated cells these membrane spanning proteins display *self* and *non-self* peptide antigens to T lymphocytes. There are two major classes of MHC molecules. MHC class I molecules

are expressed on nearly all nucleated cells. They mainly present peptides produced in the cytosol to cytotoxic CD8⁺ T cells. MHC class II proteins are exclusively expressed by professional antigen presenting cells (APCs). These cells mainly display peptides from internalized and degraded antigens in complex with MHC class II molecules on their cell surface. In humans MHC proteins are called HLA (human leukocyte antigen) and we differentiate three major class II isotypes: HLA-DP, -DQ and -DR.

Autoimmune and hypersensitivity disorders are usually multifactorial diseases with involvement of multiple genetic loci and environmental factors. For many of such diseases, genes of the HLA gene complex are strongest genetic components, and these diseases display strong associations with certain HLA variants (Horton et al. 2004). While the major predisposing genetic factors are well described for most autoimmune diseases, the environmental factors and mechanisms causing disease are often unknown. A well studied exception is celiac disease (CD), where both the major environmental and the major genetic predisposing factor are known. During the last years CD has become a valuable model disorder, helping us to understand other complex HLA associated diseases. But despite the huge advances which were made in understanding the complex pathogenesis of CD, no specific treatment supplementing the strict gluten free diet is available for patients. However, one of the most interesting targets for a novel CD treatment is the disease associated HLA-DQ2.5 (DQA1*0501, DQB1*0201) molecule.

1.1. MHC class II

MHC class II molecules are membrane bound $\alpha\beta$ -dimeric glycoproteins, expressed on the surface of professional APCs. The three main types of professional APCs are dendritic cells, B cells and macrophages. As major players of the adaptive immune system they present peptide antigens for recognition by CD4⁺ T cells (Babbitt et al. 1985). These peptides derive mainly from the exogenous pathway. By this pathway extracellular proteins or peptides are internalized into vesicles and transported through the endocytic compartment. Subsequently they are degraded by proteases deriving from lysozymic vesicles. The antigen fragments are then transported into the MHC class II compartment (MIIC), where MHC class II peptide complexes are formed. All

MHC class II molecules are assembled in the ER and subsequently chaperoned into the MIIC by invariant chain (Ii) in form of a nonameric complex (Roche et al. 1991). Ii is interacting with the peptide binding groove and prevents premature peptide binding (Cresswell 1996). Enzymes in the MIIC cleave Ii, but leave a peptide (CLIP peptide) bound to the peptide binding groove. This peptide will finally be exchanged for an antigen derived peptide with higher affinity (Cresswell 1994), in a process controlled and catalyzed by a specialized non-classic HLA protein, HLA-DM (Denzin and Cresswell 1995).

1.1.1. The MHC:peptide complex

All MHC class II molecules consist of two glycoprotein subunits. The membrane distal parts of both extracellular domains ($\alpha 1$ and $\beta 1$) are forming the peptide binding groove as an interchain dimer. The floor of the binding groove is formed by eight strands of antiparallel β -sheets (each chain is contributing 4 strands), the walls by two α -helices, one from each domain (Brown et al. 1993). The binding region of both the α - and β -chain, is rich in polymorphic residues (Stern et al. 1994). The deep groove which is formed between the two helices is open at both sides and allows to stably bind and present peptides with variable length and a relatively broad diversity (Brown et al. 1993; Madden 1995). Peptides bound to MHC adopt an extended, polypyrrolone type II conformation (Jardetzky et al. 1996; Stern et al. 1994). The MHC:peptide complex is stabilized by hydrogen bonds mainly between conserved residues of the MHC class II molecule and the peptide backbone. Additionally, the polymorphic residues within the peptide binding groove form pockets, which can accommodate the side chains of the amino acid residues in positions P1, P4, P6, P7 and P9 of the peptide, the so called anchor positions. The peptide residues in the non-anchor positions P2, P3, P5 and P8 can be recognized by T-cell receptors (TCR). Together with the MHC surface, they form the T-cell recognition site (Stern et al. 1994). The stability of the MHC:peptide complex and especially the three-dimensional structure of the T-cell binding surface (formed by the two α -helices as well as the bound peptide) define the T-cell response. T cells are highly specific and therefore even small changes in the structure of the bound peptide can result in a drastic change of the T-cell response (Bertoletti et al. 1994; Bielekova and Martin 2001).

1.1.2. The peptide binding motif of HLA-DQ2.5

The first peptide elution experiments with HLA-DQ2.5 were performed 14 years ago and led, in combination with binding experiments using truncated and substituted derivatives, to the characterization of the peptide binding motif of this molecule (Table 1). HLA-DQ2.5 can accommodate negative charges in almost all positions of the peptide binding frame (Stepniak et al. 2008), but shows strong preference for glutamate only in positions P4, P6 and P7. The ability to accommodate many negative charges together with the preference for bulky hydrophobic residues in positions P1 and P9, are the unique characteristics of the HLA-DQ2.5 molecule (Johansen et al. 1996a; van de Wal et al. 1996; Vartdal et al. 1996).

Table 1: Characteristic features of typical amino acid residues and preferred amino acids identified in important positions of the HLA-DQ2.5 peptide binding site.

| <i>Peptide binding register position</i> | <i>Characteristic features</i> | <i>Preferred amino acids</i> | <i>Reference</i> |
|--|--|--|--|
| P1 anchor position | -bulky hydrophobic residues ¹ | P, V W, F, Y, I, V, L I, L | (Stepniak et al. 2008) (Vartdal et al. 1996) (van de Wal et al. 1996) |
| P2 | -hydrogen bond involving peptide main chain amide ² | | |
| P4 anchor position | -negative charges ¹ -hydrogen bond involving peptide main chain amide ² | D, E D, E, L, V, I D, E, A | (Stepniak et al. 2008) (Vartdal et al. 1996) (van de Wal et al. 1996) |
| P6 anchor position | -negative charges ^{4,5} -hydrogen bond involving peptide main chain amide ² | D, E, A D, E, P P, A | (Stepniak et al. 2008) (Vartdal et al. 1996) (van de Wal et al. 1996) |
| P7 anchor position | -negative charges ¹ | D, E D, E D, E, W, A | (Stepniak et al. 2008) (Vartdal et al. 1996)van (van de Wal et al. 1996) |
| P8 | | P | (Stepniak et al. 2008) |
| P9 anchor position | -bulky hydrophobic residues ^{1,3} -negative charges ⁴ -hydrogen bond involving peptide main chain amide ² | F, L, E W, F, Y, I, L, V, M W, Y, A, | (Stepniak et al. 2008) (Vartdal et al. 1996) (van de Wal et al. 1996) |

¹ (Johansen et al. 1996b), ² (Bergseng et al. 2005), ³ (Quarsten et al. 1998), ⁴ (Stepniak et al. 2008), ⁵ (Vartdal et al. 1996)

The X-ray crystal structure of HLA-DQ2.5 with the DQ2- α -I-gliadin epitope QLQPFQPELPY has been solved with a resolution of 2.2 Å (Kim et al. 2004). It demonstrated that side chains of residues at positions P4, P6 and P7 can dock into the corresponding binding pockets of HLA-DQ2.5. Further it proved that glutamate in P6 is a very important anchor residue due to two principles: by participating in a wide spanning hydrogen bonding network and by charge complementation to the Lys- β 71 residue of HLA-DQ2.5. In many gluten derived antigens there are negatively charged amino acids in position P6, but also in positions P4 and/or P7. The crystal structure showed that these charges are complemented by the positive charges of Arg- β 70 or Lys- β 71.

Interestingly, most of the binding energy in the MHC:peptide complexes is calculated to originate from the hydrogen bonding network (Stern et al. 1994). Gluten derived epitopes, however, contain many proline residues - the only proteinogenic amino acid unable to donate a hydrogen bond due to its tertiary amide nitrogen. The immunodominant DQ2- α -I-gliadin epitope for example contains four proline residues. Again, the crystal structure showed how this molecule can accommodate these many proline residues and still maintain a strong hydrogen bonding network (illustrated in Fig.1). If proline is accepted in positions P1, P3, P5 and P8, the binding energy from main chain hydrogen bonding interactions in positions P-2, P-1, P2, P4, P6, P7 and P9 is retained. Interestingly, no other HLA class II molecule appears to be able to bind proline in position P1. In HLA-DQ2.5 this seems to be possible due to a deletion mutation of the α 53 residue, causing the loss of the typical hydrogen bond which is reported for many other MHC class II molecules in this position (Ghosh et al. 1995; Smith et al. 1998; Stern et al. 1994).

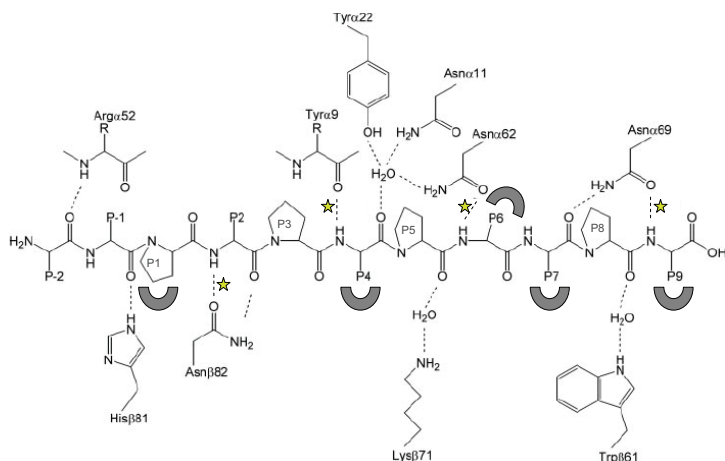


Figure 1: Overview over the complex hydrogen bonding network between conserved and polymorphic residues of HLA-DQ2.5 and the backbone amide nitrogen (-NH) and carbonyl oxygen (-C=O) of the α -I-gliadin peptide. The five pockets of HLA-DQ2.5 are indicated by grey half circles and backbone amide hydrogen bonds are shown by stars (Bergseng et al. 2005; Kim et al. 2004; Nelson and Fremont 1999).

High affinity binders are composed of favorable amino acids residues in all nine positions of the peptide binding frame. Generally, peptides with high affinity to HLA-DQ2.5 contain several negative charges (preferably in positions P4, P6 and P7), bulky hydrophobic residues in positions P1 and P9, and amino acids participating in a hydrogen bonding network in positions P2, P4, P6, P7 and P9. Important as well are additional hydrogen bonds between HLA-DQ2.5 and the peptide main chain outside of the P1-P9 core binding region (Fig. 1). Binding affinities usually increase when further amino acids are added N- and C-terminally to the core binding region (Bartnes et al. 1999; Kim et al. 2004).

1.1.3. Gluten derived peptides presented by HLA-DQ2.5

Wheat gluten and related grain storage proteins from rye and barley are the harmful molecules in CD. Gluten is a complex protein composite of gliadins and glutenins. The gliadins can be grouped into α -, γ -, and ω -gliadins (Wieser et al. 1987), a nomenclature based primary on their sequence characteristics and on their electrophoretic mobility at low pH. Glutenins are grouped into high- and low

molecular weight subunits. Both subunits contain intrachain disulfide bonds and are poorly soluble in water and other aqueous solutions. In addition, the glutenins contain interchain disulfide bonds. Gluten proteins are unusually rich in proline (~15%) and glutamine (~30%). Immunogenic epitopes which are typically presented by HLA-DQ2.5 are often found to cluster in proline-rich regions. There are several probable reasons for this epitope clustering. An important reason is that proline is protecting larger fragments from complete digestion. The human intestine is lacking enzymes feasible to degrade peptides with high proline content, the so called endoprollyl peptidases. As a consequences, large proline rich peptides survive digestion in the small intestine, and these fragments are shown to cross the mucosal surface into the lamina propria (Hausch et al. 2002). In individuals with CD, patrolling APCs with HLA-DQ2.5 and/or HLA-DQ8 on their surface will recognize special epitopes from proline rich regions and bind and present them in complex with HLA to T cells. It is also shown that gluten derived epitopes with many proline residues adopt a left-handed polyproline type II (PPII) helical conformation. This might support binding to HLA-DQ2.5, as HLA molecules are shown to cause a PPII conformation in their bound ligands (Jardetzky et al. 1996;Parrot et al. 2002).

I described earlier that HLA-DQ2.5 prefers negatively charged residues in almost all positions of the peptide binding frame. Natural ligands as well as gluten derived epitopes carry many negative charges. Gluten however is not rich in glutamate, but in non-charged glutamine residues. This paradox was solved when the enzyme tissue transglutaminase (TG2) was observed to perform deamidation of specific glutamine residues in gluten derived epitopes (Molberg et al. 1998). This deamidation is introducing negative charges which in turn increase the peptide binding affinity to HLA-DQ2.5 (Fleckenstein et al. 2002;Qiao et al. 2005). The glutamate residues which are targeted by TG2 are almost always in position P-2 of a proline residue. This specific deamidation is another probable reason why epitopes cluster in proline-rich sequences. And finally, it is observed that multiple binding sites can enhance the binding affinity of a peptide compared to single binding sites (Sette et al. 1990). The highly immunogenic 33mer gliadin peptide for example is shown to be proteolytically stable and contains multiple copies of three different T-cell epitopes and 13 proline residues (Shan et al. 2002).

1.1.4. Detergent-solubilized and water soluble MHC class II molecules

Detergent-solubilized MHC class II molecules can be purified from homozygous Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell lines. These molecules are widely used to elucidate peptide binding properties by elution and identification of natural bound ligands and in peptide binding assays to analyze binding affinities of synthetic peptides (Falk et al. 1994; Jardetzky et al. 1990; Krieger et al. 1991; O'Sullivan et al. 1990; Rammensee et al. 1995). Also the binding motif of HLA-DQ2.5 was characterized by using such detergent-solubilized molecules (Johansen et al. 1994; van de Wal et al. 1996; Vartdal et al. 1996).

Water soluble MHC molecules were made to study interactions between MHC, peptide and TCR in a detergent free environment (Wettstein et al. 1991) and in order to solve the crystal structure of these complexes (Kozono et al. 1994; Stern et al. 1994). Today, these molecules are widely used in tetramer technology, for the identification, quantification and isolation of specific T cells.

Expression of MHC class II molecules in insect cells became a commonly used method for the production of recombinant molecules (De Wall et al. 2006; Stern and Wiley 1992; Wallny et al. 1995). Initially, HLA-DR1 and other MHC class II molecules were expressed without a covalently tethered peptide as so called “empty” class II molecules (Stern and Wiley 1992). The expression of recombinant molecules with a high affinity ligand covalently tethered to the N-terminus of the β -chain (Crawford et al. 1998) however gave improved expression and more stable molecules. Therefore this strategy developed as a preferred expression modality. In our hands the “empty” HLA-DQ2.5 molecules, which we expressed in a *Drosophila melanogaster* S2 cell line, gave unusually high protein quantity. However, these molecules were not empty, instead an insect cell derived protein was stably bound to the peptide binding site (Jüse et al. 2009).

1.2. Celiac disease

CD is a gluten induced, autoimmune-like inflammatory disorder of the small intestine and therefore often referred to as gluten sensitive enteropathy. As for many diseases with autoimmune characteristics, CD is associated with certain HLA molecules. Over

90% of the CD patients express HLA-DQ2.5 and most of the remaining express HLA-DQ8. Gluten from wheat seeds and related storage proteins in seeds from barley and rye are known as the environmental triggers of CD (Dicke et al. 1953;Trier 1991;Wieser 1995;Wiesner et al. 2008). It has been shown that peptides deriving from both subgroups of gluten, the gliadins and glutenins, are toxic for CD patients (Ciclitira et al. 1984;Dewar et al. 2006;Lundin et al. 1993;Lundin et al. 1994;Molberg et al. 1997;Vader et al. 2002). Upon ingestion of these proteins, young individuals with CD often present with typical symptoms like chronic diarrhea, abdominal pain, malabsorption and failure to thrive (Maki and Collin 1997). In adult patients, symptoms are often less typical and may include weight loss and extra intestinal symptoms like neurological and behavioral disorders (Fasano and Catassi 2001). CD is diagnosed in about 0,5-1% of the Western civilization (Dube et al. 2005;Maki et al. 2003). Diagnosis is especially difficult in adults, due to the often more unspecific symptoms, silent CD or disease latency. Therefore, the actual CD prevalence in the population might be even higher then calculated (Fasano and Catassi 2001;Maki and Collin 1997). The only successful treatment today is a life-long total exclusion of gluten and related proteins from the diet. This treatment has a very good benefit-risk ratio, however, due to poor diet compliance or silent disease the risk increases for developing other autoimmune diseases (Cosnes et al. 2008;Ventura et al. 1999), diverse neurological symptoms (Cooke and Smith 1966;Zelnik et al. 2004), infertility (Sher and Mayberry 1994), osteoporosis (Molteni et al. 1990) and intestinal lymphoma (Holmes et al. 1989;Maki and Collin 1997). Therefore, there is an unmet need for alternative treatments.

1.2.1. Pathogenesis

The pathogenesis of CD is a complex interplay of the innate and adaptive immune system, which in CD patients become both activated upon gluten ingestion (Jabri and Sollid 2009). As described above, especially the proline rich fragments of gluten proteins survive gastrointestinal digestion and are able to cross the mucosal epithelium into the lamina propria. Eventually they get enzymatically deamidated by TG2 (most epitopes are dependent on this conversion), a modification which is introducing negative charges. APCs in the lamina propria can bind the now negatively charged epitopes and present them in complex with HLA-DQ2.5 or HLA-DQ8 on

their surface. These complexes are recognized by specific gluten reactive CD4⁺ T cells, which upon recognition will become activated (Molberg et al. 1998;Qiao et al. 2009). This activation step can be modulated by a secondary signal which is especially important for successful activation of naïve T cells. A secondary signal is provided by a variety of co-stimulating molecules such as CD28, CTLA-4 and CD40L on the T cell, as they interact with their counterparts on the APC surface. Following activation, naïve CD4⁺ T cells can differentiate into different functional subsets. The cytokine environment, the type of involved co-stimulating molecules and the dose of stimulation determine the kind of T-cell subset. The Th1 subset, which is associated with inflammation and tissue damage in autoimmune diseases, is the major subset found in CD (Nilsen et al. 1995). Typical for gluten reactive Th1 cells is the huge production of IFN- γ (Ben-Horin et al. 2006;Nilsen et al. 1998). As T-helpers these cells promote B-cell development into plasma cells which produce IgG and IgA antibodies directed against (deamidated) gluten and TG2 (Dieterich et al. 1997;Lerner et al. 1994;Sollid et al. 1997).

Another T-cell population typically increased in the celiac lesion is the intraepithelial lymphocytes (IEL), mostly of the TCR $\alpha\beta^+CD8^+CD4^-$ cytotoxic T-cell subtype. These T cells belong to the innate immune system. In biopsies from CD patients, up to 70-80 IEL are found per 100 duodenal epithelial cells. In patients on a GFD, a decrease in the number of IEL is only partly seen (Ferguson and Murray 1971). Therefore, IEL have become an indicator for CD. Cytotoxic IEL are also found to express the heterodimeric CD94/NKG2 receptors on their surface which are reported to have inhibitory or activating functions. In the healthy state, CD94/NKG2A⁺ TCR $\gamma\delta^+$ IEL produce the transforming growth factor (TGF)- β when they interact with their ligand HLA-E on the surface of cells of the mucosal epithelium (Bhagat et al. 2008). This is thought to suppress the activation of cytotoxic IEL. A massive IL-15 production by stressed epithelial cells however is blocking the TGF- β effect and in turn activates an up-regulation of a different NK receptor, CD94/NKG2D. This receptor allows the recognition of stressed enterocytes expressing MICA (Roberts et al. 2001). Upon recognition, the IEL is directly cytotoxic to the MICA expressing stressed epithelial cells (Hüe et al. 2004;Jabri and Sollid 2006;Meresse et al. 2006). Typically, cytolysis is performed in a TCR independent manner and without inducing IEL proliferation or cytokine production (Hüe et al. 2004;Maiuri et al. 2000;Maiuri et al. 2003). The

gliadin induced process of tissue damage is dependent on the IL-15 production and can be inhibited by antibodies neutralizing IL-15 (Maiuri et al. 2003). It is further proposed that IL-15 production is directly induced by gliadin or by the gliadin-derived peptide p31-49 (Hüe et al. 2004).

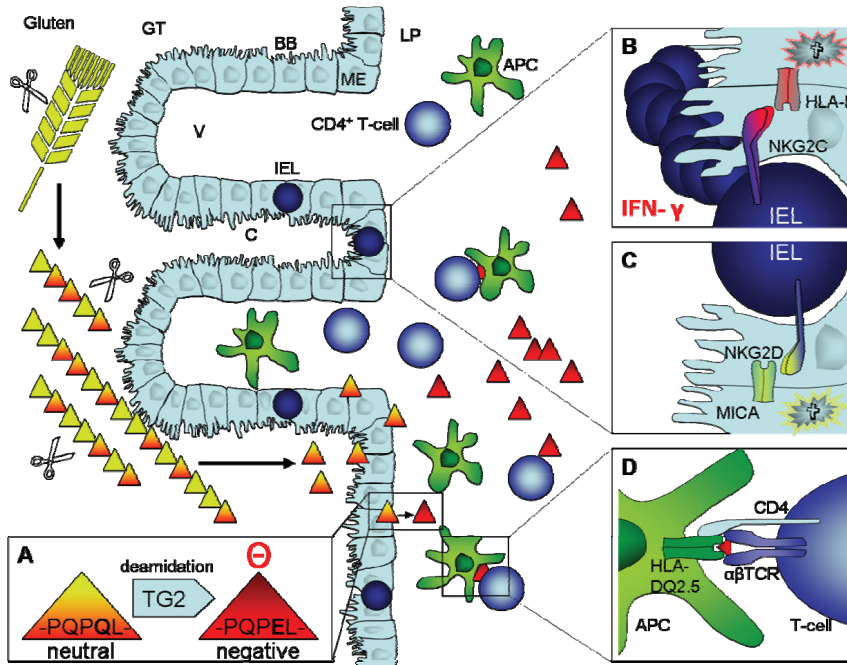


Figure 2: The pathogenesis of CD. In the gastrointestinal tract (GT), gluten is broken down to peptides. Proline rich regions (orange triangles) survive digestion and cross through the mucosal epithelium (ME) into the lamina propria (LP). The enzyme transglutaminase 2 (TG2) is introducing negative charges in certain positions (insert A). Stressed ME cells up-regulate HLA-E expression. Recognition by NKG2C on the surface of intra epithelial lymphocytes (IEL) triggers tissue damage, $\text{INF-}\gamma$ production and IEL proliferation (insert B). Also recognition of MICA, expressed on the surface of stressed ME by NKG2D, triggers tissue damage (insert C). Gluten up-regulates both, the expression of MICA and NKG2D molecules. In the LP, negatively charged, proline rich gluten derived peptides are encountered by antigen presenting cells (APCs) which bind and present them in complex with HLA-DQ2.5 molecules to specific CD4^+ T cells (insert D). This interaction will trigger T-cell activation and $\text{INF-}\gamma$ production. Further abbreviations: brush border (BB), villus (V), crypt (C).

At the same time, HLA-E molecules on epithelial cells are reported to be strongly up-regulated in untreated CD, which as well might be a stress-induced reaction (Perera et al. 2007). It is speculated that this reaction may be a consequence to IFN- γ release by IEL and CD4⁺ T cells (Guy-Grand et al. 1998; Jabri et al. 2000; Meresse et al. 2006). CD94/NKG2C positive IEL might recognize the up-regulated HLA-E and initiate further IFN- γ secretion, proliferation and tissue damage (Meresse et al. 2009).

1.2.2. Clinical diagnosis

For a defined diagnosis of CD, it is necessary to analyze a biopsy of the small intestine. Typical biopsy findings for an active CD are blunting of the villi, crypt cell hyperplasia and lymphocyte infiltrations into both the epithelium and the lamina propria. Before taking a biopsy however, a blood sample is often drawn to measure the highly specific, gluten-dependent IgA and IgG autoantibodies to TG2. Only CD patients with active disease, not healthy controls nor CD patients on a gluten free diet, have anti-TG2 IgA and IgG autoantibodies in their blood (Dieterich et al. 1998; Fasano and Catassi 2001). When a CD patient starts a strict gluten free diet, the anti-TG2 antibody titer drops, but the anti-TG2 antibodies reappear quickly upon gluten challenge. The intestinal villous atrophy and the signs of a flat mucosa take a lot more time to restore, up to several months or even years. The histological alterations during this progress are graded by the Marsh criteria, a scale ranging from Marsh III (complete villous atrophy) to Marsh 0 (normal mucosa) (Marsh et al. 1968). The original classification by Marsh has been partly modified later (Oberhuber et al. 1999).

1.2.3. Genetics

As 90-95% of the CD patients express HLA-DQ2.5 (DQA1*0501, DQB1*0201), there is no doubt that this HLA molecule is the major genetic risk factor (Sollid et al. 1989). Patients not expressing HLA-DQ2.5 are mostly HLA-DQ8 (DQA1*0301, DQB1*0302) positive (Spurkland et al. 1992). HLA-DQ2.2 (DQA1*0201, DQB1*0202) is not a risk factor and only a few HLA-DQ2.2 positive patients, most of these being homozygous, have been identified among the rare non-DQ2.5, non-DQ8 CD patients (Karell et al. 2003; Sollid and Thorsby 1993). HLA-DQ2.5 and

HLA-DQ2.2 are highly homologous and share almost identical peptide binding motifs. HLA-DQ2.2 however has been shown to have an increased off-rate for immunogenic gliadin peptides compared to HLA-DQ2.5, which might prevent successful presentation to specific T cells in the gut associated lymphoid tissue (Fallang et al. 2009). The increased stability of the HLA-DQ2.5:peptide complex is caused by an additional hydrogen bond from the polymorphic residue Tyr-22 α to the peptide main chain. In HLA-DQ2.2 this tyrosine residue is replaced by a phenylalanine (Phe-22 α) which is unable to donate this hydrogen bond.

HLA-DQ2.5 is encoded in cis (DR3-DQ2) or trans (DR7-DQ2, DR5-DQ7) (Sollid et al. 1989). Interestingly, the DR3-DQ2 haplotype is found to be associated with many different autoimmune disorders besides CD (type 1 diabetes, lupus, Graves' disease) (Price et al. 1999; Thorsby and Lie 2005).

HLA-DQ2.5 alone does however not sufficiently explain disease predisposition as also 25% of the healthy individuals express HLA-DQ2.5 (Sollid and Thorsby 1993). Also the difference between the concordance rate of monozygotic twins (70-75%) (Nistico et al. 2006) and the concordance rate of HLA-identical dizygotic twins (30%) argues for additional genetic factors (Greco et al. 2002). It has been calculated that HLA genes account for about 40-50% of the genetic predisposition (Meresse et al. 2009; Sollid and Lie 2005). But what are these additional contributing genes? A number of studies mapping non-HLA genes for CD association, thereof several genome-wide association and follow-up studies, pointed out a number of possible candidate genes, most of them with immune function (Dubois et al. 2010; Greco et al. 2001; Hunt et al. 2008; van Belzen et al. 2003; van Heel et al. 2005; van Heel et al. 2007). However, no additional high risk non-HLA gene could be identified so far.

1.2.4. Treatment possibilities today and in the future

The only treatment option today is the total exclusion of gluten and related proteins from the diet. This is for many patients a difficult task, as gluten is widely used by the food industry and potentially present in every processed food product if not declared as "gluten free". Gluten free products are available, but usually these are more expensive, and often they are not as palatable as the gluten containing products. It is

estimated that even of the diagnosed CD patients only 40-80% comply to a gluten free diet (Högberg et al. 2003;Leffler et al. 2008). Incompliant CD patients and individuals with silent and therefore undiagnosed disease are shown to have an increased risk to develop CD associated diseases (Corrao et al. 2001;Fasano and Catassi 2001).

Several new treatment strategies are currently under investigation. These belong mainly into two different classes; the modification of gluten and the specific interaction with the immune response. New wheat strains with low or absent toxic gluten proteins could be engineered, or different plant cultivars, containing storage proteins with similar properties, but without antigenic epitopes, could be used as a wheat substitute (Frisoni et al. 1995;Molberg et al. 2005;Spaenij-Dekking et al. 2005b;Spaenij-Dekking et al. 2005a). Yet, it has been a challenge to engineer modified wheat due to the large numbers of epitopes in gluten and similar proteins. Another strategy is the use of specific enzymes able to cleave proline rich epitopes. These could be supplemented to the gluten containing normal diet. Different endopeptidases have been described. Prolyl-endopeptidase (PEP) together with cysteine endoprotease (Gass et al. 2007;Gass and Khosla 2007;Shan et al. 2002;Siegel et al. 2006;Siegel et al. 2007) and the PEP from *aspergillus niger* (AN-PEP) (Stepniak et al. 2006) are currently tested in clinical trials. An alternative to this is the pre-treatment of gluten containing food with bacteria derived peptidases (Di Cagno et al. 2004). Blocking TG2 would prevent the introduction of negative charges into immunogenic gluten epitopes (Siegel and Khosla 2007). There are, however, gluten epitopes known which do not require deamidation (Dewar et al. 2006;Vader et al. 2002) and secondly TG2 has diverse biological roles which in parts remain unknown. Recently, a polymeric binder with gluten sequestering ability was presented (Pinier et al. 2009). This polymer aims to retain gluten and to prevent the genesis of smaller fragments which could cross the mucosal epithelium. A novel zonulin inhibitor (AT1001) has been developed and tested in clinical trials to reduce the gluten induced epithelial barrier dysfunction (also called “leaky gut”) (Paterson et al. 2007). A tight epithelium is thought to decrease gluten epitope exposure in the lamina propria. AT1001 was found to be well tolerated and it could improve the gastrointestinal symptoms in CD patients. Silencing of gluten specific T cells by altered peptide ligands (de Haan et al. 2002;de Haan et al. 2005;Sloan-Lancaster and Allen 1996) or soluble HLA-DQ2:peptide complexes has been tested successfully for different

epitopes, however, a therapy with such complexes is difficult due to the polyclonal nature of T cells recognizing the same antigen. An additional limiting factor for peptide vaccination is the large number of epitopes (dominant and infrequent) and the requirement of tolerance induction to all of them. However, the first clinical trials are expected to start soon (Sollid and Lundin 2009). A more exotic strategy, currently in clinical trials, is the use of an intestinal hookworm to modulate the type of the immune response (Reddy and Fried 2009). Cytokine therapy is widely used to treat different autoimmune diseases. However, due to the typical side effects of such therapies and a low acceptance for side effects in CD treatment (due to that CD is not life threatening and GFD as a safe and effective treatment is available), none of the known cytokine therapy options are of interest.

After all, due to a strong HLA-DQ2.5 association (over 90% of the CD patients express HLA-DQ2.5 (DQA1*0501, DQB1*0201)), this molecule is likely the most interesting target for the development of a specific treatment for CD and other HLA-DQ2.5 associated diseases.

1.2.5. Blocking presentation and/or recognition of gluten derived peptides

Peptides or peptide-like molecules (peptidomimetics) with high affinity to HLA-DQ2.5 could be used to interfere with the presentation of gluten derived immunogenic peptides. So far, the most promising approaches for blocking the HLA-DQ2.5 mediated presentation of gluten derived antigens is the use of cyclic and dimeric gluten peptides (Xia et al. 2007), or the introduction of a large, chemically modified side chain into gluten epitopes (Kapoerchan et al. 2008). However, the efficacy of the blockers tested so far in *in vitro* T-cell assays is limited, probably due to insufficient affinity of the blockers to HLA-DQ2.5. In order to further develop this approach, the binding affinity of a HLA-DQ2.5 blocker has to be several orders of magnitude higher than the affinities of gluten derived antigens. In addition, blockers used for CD treatment in humans should fulfill several criteria. These substances should be safe, inexpensive, administered orally and should not introduce a potential risk for T-cell sensitization.

2. Aims of this study

The major aim of my PhD project was to study peptide binding to HLA-DQ2.5 by means of peptide library based methods. The main focus herein was to establish a novel peptide library based method using recombinant, water soluble HLA-DQ2.5 molecules and soluble complex peptide libraries.

The specific aims towards that goal were:

- I. Design and synthesis of different soluble, synthetic peptide libraries.
- II. Production of suitable soluble, recombinant HLA-DQ2.5 molecules.
- III. Establishment of a method to isolate high affinity HLA-DQ2.5 binders from complex peptide mixtures.
- IV. Identification of the isolated high affinity peptides by tandem mass spectrometry and bioinformatics tools.

With this novel peptide library approach, we can select for peptides with the highest HLA-DQ2.5 affinity from complex mixtures of synthetic peptides. For blocker development, also modified non-natural amino acids could be incorporated within these peptide libraries.

In a different approach, we used a positional scanning peptide library and designed high affinity binders by combining only favourable amino acids in each position of the binding frame. The sequence with highest HLA-DQ2.5 affinity from these experiments is an interesting lead compound for the development of HLA-DQ2.5 blockers.

3. Summary of the papers

Paper 1:

Soluble HLA-DQ2 expressed in S2 cells copurifies with a high affinity insect cell derived protein

Ulrike Jüse, Burkhard Fleckenstein, Elin Bergseng, Ludvig M. Sollid

Immunogenetics 2009 Feb;61(2):81-9

In this paper we characterize the interaction between the recombinant soluble HLA-DQ2.5 (sDQ2) molecule and the *Drosophila melanogaster* derived protein DCB-45. This work was motivated by the stable complex formation and our interest in the development of reagents which inhibit HLA-DQ2.5 peptide binding.

Paper 2:

Design of new high-affinity peptide ligands for human leukocyte antigen-DQ2 using a positional scanning peptide library

Ulrike Jüse, Yvonne van de Wal, Frits Koning, Ludvig M. Sollid, Burkhard Fleckenstein

Hum Immunol. 2010 May;71(5):475-81

A positional scanning peptide library approach was used to create high affinity binders to HLA-DQ2.5. The peptide with highest affinity bound 50 fold stronger than the immunodominant DQ2- α -I-gliadin epitope and could serve as a lead compound in HLA-DQ2.5 blocker development.

Paper 3 - manuscript

Assessing high affinity binding to HLA-DQ2.5 by a novel library-based approach

Ulrike Jüse, Magnus Arntzen, Peter Højrup, Burkhard Fleckenstein and Ludvig M. Sollid

Manuscript (submitted to Bioorganic & Medical Chemistry, 03/2010)

In this manuscript we describe a novel peptide library based method which can be used to identify optimal binders for HLA class II molecules. In short, complex

synthetic peptide libraries were incubated with recombinant soluble HLA-DQ2.5 molecules. Selected library peptides were eluted and analyzed by tandem mass spectrometry online coupled to a nano-liquid chromatography. Subsequently, peptide sequences were identified using the Mascot search engine and manual evaluation.

4. Methodological considerations

4.1. Peptide library approach

In my PhD project we have developed a novel peptide library based method, using soluble complex peptide libraries with one or several randomized positions (X) and soluble recombinant HLA-DQ2.5 molecules (sDQ2). Figure 3 is illustrating schematically a peptide library with 6X positions which is interacting with the HLA-DQ2.5 binding groove.

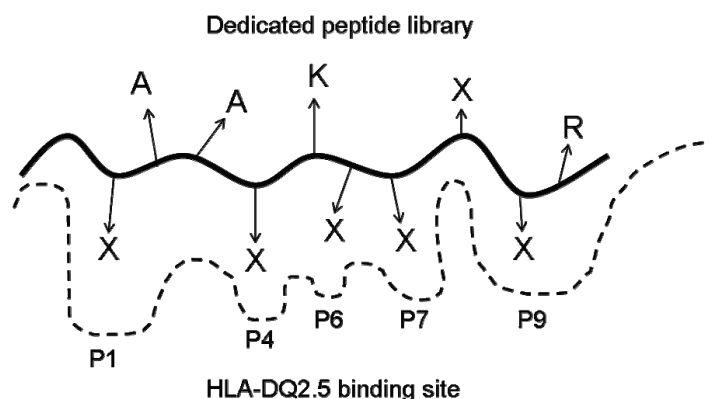


Figure 3: Schematic representation of a peptide library which is interacting with the peptide binding groove of HLA-DQ2.5. The residues in the positions P1, P4, P6, P7 and P9 can interact as anchors with the pockets of the binding site. The residues in the positions P2, P3, P5 and P8 are solvent exposed. X is symbolizing dedicated positions in which an equimolar representation of 2-13 selected amino acids was chosen (the composition in the X positions is defined in table 2).

By incubating sDQ2 with a large excess of a complex peptide mixture, we aimed to create an environment where only the best binders are able to compete for limited sDQ2 binding sites. Peptides bound to sDQ2 can then be isolated by preparative size exclusion chromatography and subsequently analyzed by LC-MS/MS. In the peptide libraries the sequence length, the alignment of binding registers, and the numbers and composition of randomized positions can be chosen and non-natural amino acids can

be included. Peptide libraries with many randomized positions are highly complex and can be used for the identification of optimal binders. Libraries which contain a single randomized position can be used to study the effect of single residue substitution.

Each of our peptide libraries are composed of randomized (fully randomized or dedicated) and defined positions (Table 2).

Table 2: Peptide library formats used for the identification of high affinity binders. Amino acids are given in the one letter code. X in the libraries I-IV encodes an equimolar mixture of all natural amino acids, except for cysteine. In the libraries V-VIII, X represents a dedicated position with a composition as listed. The position of each amino acid in the peptide binding frame is indicated by the subscribed number.

| peptide binding register | |
|--------------------------|---|
| I | G-G- F ₁ -A ₂ -P ₃ -E ₄ -K ₅ -E ₆ -E ₇ -P ₈ -X ₉ - R-NH ₂ |
| II | G-G- X ₁ -A ₂ -P ₃ -E ₄ -K ₅ -E ₆ -E ₇ -P ₈ -F ₉ - R-NH ₂ |
| III | G-G- X ₁ -A ₂ -P ₃ -E ₄ -K ₅ -E ₆ -E ₇ -P ₈ -X ₉ - R-NH ₂ |
| IV | G-G- F ₁ -A ₂ -P ₃ -E ₄ -K ₅ -E ₆ -E ₇ -X ₈ -X ₉ - R-NH ₂ |
| V | G-G- F ₁ -A ₂ -P ₃ -E ₄ -K ₅ -X ₆ -X ₇ -P ₈ -X ₉ R-NH ₂ |
| VI | G-G- X ₁ -A ₂ -P ₃ -X ₄ -K ₅ -X ₆ -X ₇ -P ₈ -X ₉ R-NH ₂ |
| VII | Ac- A ₁ -A ₂ -A ₃ -A ₄ -K ₅ -X ₆ -X ₇ -X ₈ -X ₉ R-NH ₂ |
| VIII | Ac- X ₁ -A ₂ -A ₃ -X ₄ -K ₅ -X ₆ -X ₇ -X ₈ -X ₉ R-NH ₂ |

Ac: acetylated N-terminus; -NH₂: amidated C-terminus.

Amino acid composition in the X positions of dedicated libraries V-VIII:

X₁: P, V, E, W, F

X₄: P, V, T, G, W, L, Y, N, E

X₆: Q, T, N, Y, F, V, G, L, A, P, E

X₇: G, L, F, A, E

X₈: P, E

X₉: G, Q, A, S, V, P, L, E, Y, D, M, F, W

The positively charged lysine at position P5 and proline in position P3 or P8 are defined amino acids. The side chains of amino acids in these three positions point away from the binding groove (Fig. 3) and there are no main chain hydrogen bonds in position P3 and P8 (Fig. 1). We used these amino acids in non-anchor positions to ensure that all library peptides bind in a uniform register. If the binding frame is

shifted by one position, the positively charged side chain of lysine would point into the positively charged pocket of the binding groove. In addition, essential main chain amide nitrogen hydrogen bonds would be abolished by proline in any of the positions P2, P4, P6, P7 and P9.

Alanine is another defined amino acid we used in our libraries. The side chain of alanine is too short to be expected to act as a prominent anchor residue, at least in pockets which are deep. Library VII in table 2 for example is completely anchorless at the N-terminal part of the sequence; this will in theory focus the selection pressure on the four C-terminal X positions.

Libraries I-VI are extended by glycine residues in order to allow formation of hydrogen bonds in positions P-1 and P-2 (Fig. 1). It has been shown that additional hydrogen bonds N-terminally of the peptide binding frame increase the binding affinity to HLA-DQ2.5 (Bartnes et al. 1999;McFarland et al. 1999;Nelson and Fremont 1999). Libraries VII and VIII are acetylated in position P1. N-terminal acetylation instead of extension is minimizing the risk for register shifting, but retains the possibility to generate one of the hydrogen bonds N-terminally of the binding frame.

The C-terminal arginine (R) in position P10 of all of our peptide libraries is included to prevent shifting and to enhance signal intensities of peptides and their y-fragments in mass spectrometry.

In libraries I-IV, X is denoting an equimolar distribution of 19 natural L-amino acids (excluding cysteine). Peptide libraries with many fully randomized X positions are too complex and difficult to analyze, therefore we used dedicated X positions in the libraries V-VIII.

4.2. Synthesis of peptides and peptide libraries

For this work, peptide libraries as well as single peptides were synthesized. Peptide libraries were used to identify peptides with highest affinity from a mixture of

sequences. Individual peptides were used in peptide binding experiments to determine relative binding affinities of the identified ligands.

All single peptides and peptide libraries I-IV were synthesized on a pipetting robot (Syro I) using a standard Fmoc (Fluorenylmethoxycarbonyl)/HBTU (2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) strategy on solid supports (mainly Rink amide MBHA-resin). Randomized library positions (X) were introduced by two coupling steps using amino acid mixtures in an equimolar ratio with respect to the coupling sites of the resin. The identity of the single peptides was confirmed by MALDI-TOF MS, and the purity was determined by RP-HPLC. Some single peptides were acetylated at their N-terminus, others were labeled with a fluorophore or a radio label (^{125}I). Such labeled peptides were used as reporters in peptide binding assays, such as the HLA-DQ2.5 high affinity ligand P198 (KPLLIIAEDVEGEY, *Mycobacterium bovis* 65kDa Hsp 243-255Y). The radio-labeling with ^{125}I was done with the chloramine T method (Greenwood et al. 1963).

4.3. HLA-DQ2.5 as full length molecule or soluble construct

Different HLA-DQ2.5 reagents were used for various experiments. Besides detergent-solubilized full length molecules, we used recombinant water soluble HLA-DQ2.5 constructs.

For peptide binding assays with the radioactively labeled reporter peptide, we used detergent-solubilized HLA-DQ2.5 (DQA1*0501, DQB1*0201). These full length molecules were derived from Epstein-Barr virus-transformed B lymphoblastoid cell lines (Johansen et al. 1994).

For identification of optimal binders from the peptide libraries and for measuring HLA-DQ2.5 binding affinities of high affinity binders, we chose a water soluble recombinant construct carrying an α -I-gliadin peptide covalently attached to the β -chain by a thrombin cleavable linker (sDQ2- α I). These molecules were produced in high purity by baculovirus transfected Sf9 insect cells, which can be grown in spinner cell cultures (Quarsten et al. 2001).

Water soluble HLA-DQ2.5 without a tethered peptide (sDQ2) was expressed in a stably transfected *Drosophila melanogaster* S2 cell line. In these constructs the extracellular parts of the DQ α - and DQ β -chain are fused with a Fos- and a Jun zipper, respectively (Paulsen et al. unpublished). The S2 cells were also grown in cell spin flasks at 22°C in serum free media. The secretion of sDQ2 into the medium was induced by CuSO₄. As described in paper I, these molecules co-purified with the insect cell derived protein DCB-45.

All of our HLA-DQ2.5 molecules were affinity purified from the culture media or cell lysate using the specific, monoclonal antibody 2.12.E11 bound and cross-linked to protein A-sepharose (Johansen et al. 1994; Viken et al. 1995). Protein concentrations were measured using a BCA kit and purity was checked by SDS-PAGE.

4.4. Isolation and identification of high affinity HLA ligands from peptide libraries

For this approach we used water soluble HLA class II molecules and soluble dedicated peptide libraries. A limited amount of thrombin treated sDQ2- α I was incubated with a complex peptide library composed of a large number of different sequences in a defined format, together with soluble HLA-DM. After two days of incubation at 37°C and pH 5.3, the library peptides bound to sDQ2 were isolated from the excess of free peptides by preparative size exclusion chromatography. Then, the protein fraction was concentrated and the pH was adjusted to 1 in order to release bound peptides. The eluted peptide fraction was purified and desalted using a reverse phase material. Samples were analyzed by MALDI-TOF MS using an Ultraflex II instrument or by nLC-MS/MS using an LTQ-Orbitrap XL mass spectrometer. Peptide libraries with one fully randomized X position (libraries I and II in table 2) were analyzed by MALDI-TOF MS and the peptide sequences were identified manually from the acquired mass spectra. More complex libraries (libraries III-VIII) were analyzed with nLC-MS/MS and data were searched against a library defined database by using the Proteome Discoverer software and the Mascot search engine.

The format of the more complex peptide libraries allows for peptides with isobaric amino acid combinations. Such isobaric peptides share most of their fragment ion

masses. Therefore a manual inspection of all MS/MS events for which Mascot assigned several isobaric peptides as significant was necessary. The manual inspection included a comparison of the observed fragmentation pattern with the theoretical fragmentation of all assigned hits. Further, since isobaric peptides often elute at slightly different time points, we could compare the elution profiles and retention times for most of the isobaric peptides in the sample selected by HLA-DQ2.5 with the corresponding elution profile and retention times of peptides in the original library sample.

4.5. Peptide binding assays

To assess the relative binding affinity of single peptides, we used two different peptide binding assays. In both assays we measured the ability of the test peptide to compete with a high affinity indicator peptide for HLA-DQ2.5 binding. The binding affinity of a peptide is usually defined by its IC₅₀ value, which is the concentration of this peptide needed to inhibit 50% of the binding of the indicator peptide.

In the binding assay with the radioactively labeled P198 indicator peptide, we used detergent solubilized full length HLA-DQ2.5, a simple spin column separation method and a gamma-counter (Buus et al. 1995;Johansen et al. 1994). IC₅₀ values were calculated from the radioactivity bound to HLA-DQ2.5 (in the void volume) and the unbound radioactivity (radioactive indicator peptide on the column).

For measuring IC₅₀ values of peptides with high affinity, sDQ2- α I was used together with a carboxyfluorescein labeled P198 indicator peptide. After incubation, the protein fraction was separated from the unbound peptides by size exclusion chromatography. The UV (214nm) and the fluorescence signal (Ex. 490nm, Em. 520nm) were monitored. IC₅₀ values were calculated on the basis of the heights of the fluorescence signal to the corresponding UV signal.

Notably, IC₅₀ values derived from these two assays can not be directly compared to each other. One reason is the difference in detection sensitivity. Detection of a fluorescent indicator is less sensitive compared to the detection of radioactivity. Therefore, the fluorescent indicator peptide is used in a higher concentration, which in

turn requires more competitor peptide to achieve a 50% inhibition. Another reason might be the slower peptide exchange for sDQ2- α I which is probably caused by the slower off-rate of the DQ2- α I-gliadin epitope compared to the diverse repertoire of ligands bound to detergent solubilized HLA-DQ2.5. We tried to adjust for this by allowing peptide exchange over 48 hours instead of 12 hours.

To be able to compare results from two different experiments, we used several reference peptides in each assay. We then calculated relative binding capacities (RBC) for each test peptide by dividing the IC₅₀ values of the test peptide by the IC₅₀ value of the reference peptide P198.

5. Discussion

5.1. Interfering with the presentation of immunogenic epitopes by HLA

Most autoimmune diseases are associated with HLA molecules. HLA-DQ2.5 (DQA1*0501, DQB1*0201) is part of the B8-DR3-DQ2 haplotype and strikingly, this haplotype is associated with particularly many diseases, including CD, type 1 diabetes, systemic lupus erythematosus and Addison's disease (Sollid et al. 1989; Thorsby 1997; Thorsby and Lie 2005; Todd et al. 1987). In CD, HLA-DQ2.5 on the surface of APCs plays a key role in the pathogenesis by presenting gluten peptides to specific CD4⁺ T cells in the intestinal mucosa (Lundin et al. 1993). More than 90% of the CD patients express HLA-DQ2.5 (Sollid 2002; Sollid and Thorsby 1993). If an agent could specifically and effectively interfere with antigen presentation by this key molecule, it would allow for a novel treatment strategy for CD and other HLA-DQ2.5 associated diseases.

Blocking the presentation and/or recognition of immunogenic gluten derived epitopes can be done in two different ways. So far, the most promising approach is the use of cyclic and dimeric gluten peptides which sterically hinder the interaction of the TCR with the HLA-DQ2.5:peptide complex. Thereby the activation of the specific CD4⁺ T cell is blocked (Xia et al. 2007). Another possible approach is the blocking of the HLA-DQ2.5 binding groove by using high affinity binders to this molecule, which can effectively out-compete gluten derived epitopes. Such a HLA blocker should however not be immunogenic by itself. Therefore, it would make sense to combine both approaches by designing bulky or cyclic peptides with high affinity to HLA-DQ2.5.

Although antagonising T-cell receptors is about 100 fold more efficient than blocking antigen presentation by MHC molecules (De Magistris et al. 1992), in CD this principle is unsuitable due to the TCR heterogeneity (i.e. different TCR recognizing the same peptide-MHC complex) as well as the fairly large number of different gluten T cell epitopes. Blocking HLA-DQ2.5 might be the only way to specifically interfere with antigen presentation. For this purpose, specific high affinity binding molecules

have to be created, feasible to block the HLA-DQ2.5 binding groove for gluten derived antigens.

Since professional APCs express up to several hundred thousand MHC molecules on their surface (Stevanovic and Schild 1999; Sykulev et al. 1996) and because as little as 0.01%-0.1% of MHC:peptide complexes are needed to elicit a T-cell response, virtually a blockage of 100% has to be reached with MHC blockers to prevent T-cell activation (Christinck et al. 1991; Demotz et al. 1990; Harding and Unanue 1990). Therefore, the binding affinity of a peptide blocker for HLA-DQ2.5 should be several magnitudes higher than the binding affinities of gluten derived antigens.

Presumably, no combination of natural amino acids will fulfil such requirements for blockers. Therefore, the next essential step in blocker development will be to explore if non-natural amino acids can increase the peptide binding affinity to HLA-DQ2.5. Such modified amino acids have to be used at certain positions to optimize fitting and subsequently increase binding affinity.

5.2. Methods for identification of HLA binding motifs and optimal binders

Methods to identify peptide binding motifs and optimal ligands to HLA molecules are important in basic and applied research. Identification of a binding motif helps to understand underlying mechanisms of HLA associated diseases and to identify potential T-cell epitopes in protein sequences. The identification of optimal ligands is fundamental for the design of new drugs and vaccines as treatment for HLA associated diseases.

A variety of different techniques and methods have been used to elucidate peptide binding. In the traditional approaches, membrane bound HLA molecules are isolated from EBV-transformed B lymphoblastoid cells and the natural ligands of these molecules are eluted and sequenced. Further, these peptides are synthesized and their binding affinity is measured. By systematic truncation and single amino acid substitution experiments, the binding frames are identified and rules governing binding are established. In other approaches for HLA epitope identification, overlapping peptides covering the total sequence of a candidate protein are

synthesized, followed by direct or indirect binding assays. A different approach is the use of bacteriophages which can display various peptides (M13 phage display). These peptides can be overlapping sequences of a candidate protein or represent randomized peptide libraries (Hammer et al. 1992). Peptide displaying phages are incubated with the HLA molecule of interest. Those phages which bind to HLA are isolated and amplified, and subsequently the sequence of the peptide ligand is determined by sequencing the corresponding phage DNA.

For the investigation of high affinity binding and the potential benefits of non-natural amino acids however, peptide library based approaches provide several obvious advantages compared to the approaches described above. The format of synthetic peptide libraries can be adapted to the aim of the study. This is done by customizing the sequence length of the library, the alignment of binding registers and the numbers and composition of randomized positions. In addition, libraries of chemically modified amino acids can be included in any position.

In paper II and III we present two different library based approaches which we used to explore high affinity binding to HLA-DQ2.5. First, by a positional scanning peptide library approach the solely influence of each natural amino acid in each position of the peptide binding register was measured in an otherwise fully randomized backbone. All amino acids which were measured in each position of the peptide binding frame could be ranked by their relative contribution for binding and sorted into categories of favorable and unfavorable amino acids. In paper II we show that by simply combining the most favorable amino acids we were able to design high affinity binders. The 10mer composed of the best amino acids in each position was shown to be a 50-fold better binder to HLA-DQ2.5 than the immunodominant DQ2- α -I-gliadin peptide. This peptide was the best binder to HLA-DQ2.5 that we ever tested.

To further explore the rules for high affinity binding, we developed a novel peptide library based method (paper III), allowing us to search for optimal binders in complex peptide libraries with one or two fully randomized or several dedicated X positions. This new method also allows us to include non-natural amino acids in any position of the peptide binding frame.

5.3. Optimizing fitting to increase binding affinity to HLA-DQ2.5

5.3.1. Optimizing anchors

The crystal structure of HLA-DQ2.5 with the immunogenic gluten derived DQ2- α -gliadin epitope (QLQPFQPELPY) was solved with a resolution of 2.2 Å (Kim et al. 2004). Peptide binding is mediated by a strong hydrogen bonding network, composed of 13 hydrogen bonds from the peptide main chain to HLA-DQ2.5. Binding energy is also contributed by the fitting of amino acid side chains into the five pockets in the peptide binding groove (P1, P4, P6, P7 and P9) (Fig. 4A). Notably, the overall charge of the binding groove is positive, especially in the area of the pockets P4, P6 and P7, which explains the bias for negatively charged anchor residues found in these positions in gluten derived epitopes as well as in other ligands.

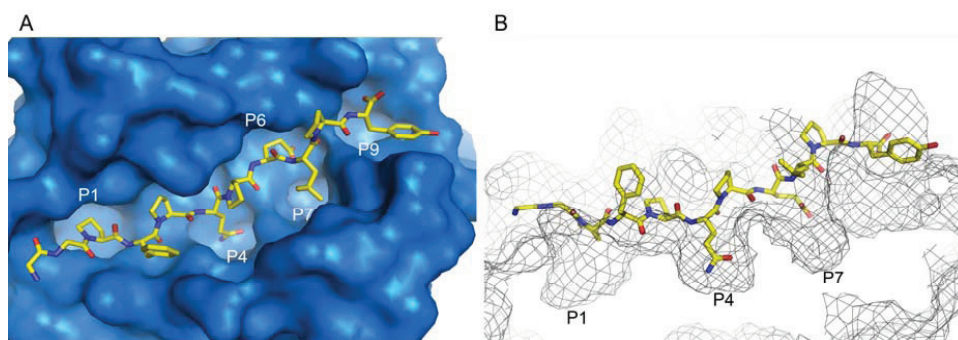


Figure 4: HLA-DQ2.5 in complex with the immunogenic gluten derived DQ2- α -gliadin epitope (QLQPFQPELPY).

- (A) Looking from above into the peptide binding groove of the HLA-DQ2.5 molecule. The surface of HLA-DQ2.5 is given in blue. The peptide is shown in yellow, with nitrogen marked blue and oxygen marked red. The pockets P1, P4, P6, P7 and P9 are labelled.
- (B) A side view of the peptide binding groove, focusing on peptide interaction with the pockets P1, P4 and P7. The HLA-DQ2.5 surface is shown as mesh, the peptide is yellow, with nitrogen marked blue and oxygen marked red.

Figures were generated with PyMol (DeLano Scientific LLC) and by using the atomic coordinates of HLA-DQ2- α -I-gliadin (PDB ID code 1S9V).

Interestingly, the closer view into the HLA-DQ2.5 peptide binding groove occupied by the DQ2- α -I-gliadin reveals an imperfect fitting of the peptide side chains in some of the anchor positions. The anchoring residues in positions P1 and P7 interact only with the surface of the deep pockets (Fig. 4A and B) and the side chain in position P9 is positioned outside, contacting the so called P10 shelf (Fig 4A). In a study by Kapoerchan et al., molecular modelling of the high affinity binder ADAYDYESEELFAA into the binding groove of HLA-DQ2.5 indicated that none of the natural amino acids are able to optimally occupy the large pocket P1 (Kapoerchan et al. 2010). Subsequently, the authors tested eight non-natural amino acids as anchor residues in position P1. Results from binding assays showed that binding affinities were increased of up to 2 fold.

In experiments using a HLA class I molecule (HLA B*2705, linked to ankylosing spondylitis), the authors reported an improved stability of the HLA-ligand complex after filling a hydrophobic binding pocket with a large non-natural aromatic side chain (Krebs et al. 1998).

These experiments were the first to successfully explore high affinity binding by using modified amino acids. However, the systematic substitution with modified amino acids is a time consuming approach since only a few residues can be tested at the same time. With a peptide library format such as the one presented in paper III, the identification of more optimal anchor residues could be done in a faster high-throughput fashion.

5.3.2. Influence of non-anchor residues

Residues in non-anchoring positions can also contribute to the overall affinity of a peptide to HLA class II molecules. In our experiments with a positional scanning peptide library (paper II) several residues in each of the non-anchoring positions showed a positive or negative effect on peptide binding. Several previous reports discuss effects of those non-anchor residues. Position P3 is one of the four non-anchor positions, but it has been referred to as a shelf with peculiar selectivity in several class II molecules: HLA-DR1 and HLA-DR3 (Ghosh et al. 1995; Stern et al. 1994), HLA-DQ6.4 and HLA-DQ6.2 (Ettinger et al. 2006) and HLA-DQ2.2 (van de Wal et al.

1997). In position P3, the side chain of the peptide has been shown not to point directly outward in the direction of the TCR like the side chains in the other non-anchor positions P2, P5 and P8. For example, the methionine of the CLIP peptide in the binding groove of HLA-DR3 is only 30% solvent exposed (Ghosh et al. 1995). In some reports on the non-disease-associated HLA-DQ2.2 molecule, the peptide residue in the position P3 is even specifically involved in peptide binding and therefore referred to as an additional anchor (van de Wal et al. 1997). Binding experiments have shown that HLA-DQ2.2 does not accept proline, histidine, arginine and lysine in this position. The disease associated HLA-DQ2.5 surprisingly does not seem to use an auxiliary P3 pocket (Ghosh et al. 1995; van de Wal et al. 1997). Indeed, the results from the positional scanning peptide library study (paper II) illustrate that position P3 in HLA-DQ2.5 is a typical non-anchor position. None of the amino acids measured in this position reached a high *relC* value. There is no particular positive effect of any amino acid in any of the four non-anchor positions, but nevertheless, some residues were found which slightly increased the overall binding affinity of the peptide sub-library, whilst others decreased binding affinities. Interestingly, only in the non-anchor positions *relC* values drop as far as -2.9. This might give some indication on the negative effect of non fitting amino acids in these positions.

To conclude, high affinity HLA class II ligands should not only have five optimal anchor residues, but also optimized residues in non-anchoring positions in order to increase the overall binding affinity. In the approach presented in paper II, we describe high affinity binders which were designed from the most favorable amino acids in each of the nine positions of the peptide binding frame.

5.4. Advantages of a water soluble approach

Screening experiments with peptide libraries are often performed on solid phase supports with the peptide bound via a linker to a resin (Lam et al. 1991). The main advantage related to this method is the rather easy detection, isolation and identification of bead-bound peptides (Lam et al. 2003). However, with peptides and proteins in solution, binding kinetics is much faster and there is no need for linkers and detachment from resin. Furthermore, unspecific binding, e.g. due to hydrophobic interactions with the solid phase, is reduced. Soluble peptide libraries can be used in

higher concentrations and/or smaller volume compared to bead based library approaches of the *one-bead one-peptide* format. The numbers of unique sequences present in the incubation mixture can be increased tremendously, giving a stronger selective force in the binding to a limited amount of HLA molecules. In addition, we use water soluble HLA-DQ2.5 molecules which allow us to work without detergent.

5.5. Limitations of the soluble peptide library approach applied

5.5.1. Sliding of the binding frame

The libraries I-VI in table 2 are N-terminally extended by two glycine residues in order to allow formation of hydrogen bonds in the positions P-1 and P-2 (Fig. 1). It has been shown that these additional hydrogen bonds increase the binding affinity of a peptide to HLA-DQ2.5 (McFarland et al. 1999). However, the most abundant amino acid we identified in the X positions of the samples eluted from HLA-DQ2.5 after incubation with libraries V and VI did not fit into the known peptide binding motif. The skewed distribution of the most abundant amino acids is probably caused by sliding of some of the sequences into an alternative frame, which is exemplified for library VI in table 3.

| frame | Relative positions | | | | | | | | | | | | | |
|-------|--------------------|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|-----|-----|
| | -P2 | -P1 | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 |
| 1 | G | G | X | A | P | X | K | X | X | P | X | R | | |
| 2 | | | G | G | X | A | P | X | K | X | X | P | X | R |

Table 3: Peptides of library VI registering in two different binding frames. Residues within the 9mer core region of the peptide binding site are given in bold. Amino acids are given in the one letter code. X represents dedicated positions (composition of each X position is defined in table 2).

Efforts to circumvent sliding included the placement of proline in position P3 and P8 and lysine in position P5. We expected that these two amino acids would force all sequences in binding frame 1 (Table 3). Proline is a special amino acid, as it can not participate in the hydrogen bonding network of the peptide main chain and therefore it is not preferred in the positions P2, P4, P6, P7 and P9. Lysine, a positively charged amino acid, is not favored in any position of the binding groove, but it is accepted in

position P5 due to its orientation towards the solvent. After sliding into binding frame 2, both proline residues have moved into positions where no hydrogen bonds to the main chain nitrogen are formed. Lysine, however, is moved into position P7. In this position, lysine is expected to introduce a penalty, but eventually it will be accepted if all other anchor residues in the bound peptide are optimal. Thus, introducing of selected amino acids into certain positions of the peptide binding frame does not guarantee that all peptides bind in the same register. N-terminal extension should be avoided and instead acetylated libraries should be used.

5.5.2. Libraries with too high complexity

A fully randomized nonapeptide library is composed of 19^9 ($3.23 \cdot 10^{11}$) unique sequences. In 1 mg of such a peptide library, each sequence is theoretically present with about 0.003 fmol, which is not enough for sequence identification. In our peptide libraries we decreased complexity by using dedicated X positions mainly placed in anchor positions. Such dedicated positions carried a reduced number of selected amino acids. Library VIII with 6 X positions is the dedicated library with the highest complexity and it contains 64 350 unique peptide sequences. When we analyzed 0.4 μ g of the original library on an LTQ-Orbitrap XL mass spectrometer online coupled to a nano-LC system, a Mascot search of the obtained LC-MS/MS data using a peptide library-defined database identified only 3% of the theoretical peptides, even though each peptide was theoretically present in pico-mol amounts (paper III). The peptides in our libraries have very similar sequences and therefore similar physicochemical properties. One dimensional LC separation might therefore be insufficient as many peptides show overlapping elution profiles. Thus, the Orbitrap will primarily select parent ions with high intensities for fragmentation, limiting the total number of identified peptides. In addition, our peptide libraries contain many isobaric peptides. These are peptides with identical molecular weight but different sequences. As co-eluting isobaric peptides will be fragmented together, they produce overlaying fragmentation patterns and therefore complicate sequence identification by Mascot and manual data inspection.

After incubation with HLA-DQ2.5 and elution of bound library peptides, the complexity of the samples is dramatically reduced and little co-elution is thus

expected. However, the Mascot search engine which we used to analyze our data showed limitations when analyzing complex synthetic peptide mixtures composed of very similar sequences. This program is designed and optimized for protein identification (Perkins et al. 1999). In our dataset, Mascot is assigning all isobaric peptide variants of an identified parent mass as significant, because they share most of their y- and b-fragments. Such events have to be inspected manually, in order to exclude potential false positive hits. In experiments with highly complex libraries, this step of manual inspection can be very time consuming.

5.6. Further optimization of the soluble peptide library approach

Our soluble peptide library approach can easily be adapted to study high affinity binders containing non-natural amino acids. I would suggest the following workflow: In a first step, nine 1X sub-libraries should be synthesized using a complex mixture of non-natural amino acids in each X position. The amino acid composition of each X position could be adapted to the characteristic features of the corresponding anchor or non-anchor position. After selection by HLA-DQ2.5, the most optimal amino acids in each position will be identified. In a second step, these optimal residues could be combined to new libraries with up to nine dedicated X positions to identify the optimal combination of non-natural amino acids. Mascot can handle non-natural amino acids and manual inspection is not required provided that no isobaric peptides are created in the dedicated libraries.

6. Conclusion

MHC molecules have developed to present a large collection of intra- and extracellular protein fragments. The limited set of different MHC molecules that one individual expresses must be able to display an universe of different antigens for T-cell recognition. Nevertheless, the peptides each molecule is presenting fit a more or less restricted peptide binding motif. MHC class I molecules are thought to be able to bind up to 10,000 different sequences (Stevanovic and Schild 1999). According to the peptide binding motif of HLA-DQ2.5 presented in table 1, it is theoretically possible to design 788,000,000 sequences with at least moderate affinity to this molecule, simply by combining the favorable residues in anchor positions with 19 natural amino acids in each of the non-anchor positions.

Since many HLA molecules are associated with autoimmunity, efforts are made to identify peptide binding motifs and epitopes. Further, peptide blockers and antagonists which can interfere specifically with the presentation of immunogenic peptides are interesting treatment options for diseases with strong HLA association. CD is a model disease in this research field. The HLA-DQ2.5 molecule is strongly associated with CD and presents various gluten fragments to a variety of gluten specific HLA-DQ2.5 restricted CD4⁺ T cells. One peptide blocker to this molecule would treat all CD patients (DQ2.5⁺; DQ8⁻), independently of their CD4⁺ T-cell specificity.

But do we need new treatment options for CD? The currently available treatment for CD is a gluten free diet (GFD). The benefit:risk ratio of the GFD treatment is very good. Unfortunately, as gluten is a widely used additive in the food industry, it is potentially present in all industrial processed foodstuffs. Gluten free food usually is expensive, often it does not taste as good and it is not available everywhere. There are many reasons why patients find it challenging to stick to their diet and for those with little or no gastrointestinal symptoms it is even harder.

A highly efficient HLA-DQ2.5 blocker is thought to allow a very specific treatment for CD. Optimally, such a blocker should be administered orally. For this application

route, it must be stable to digestion and must cross the intestinal epithelium for blocking all HLA-DQ2.5 molecules in the lamina propria. High affinity binding peptides with several proline residues or chemically modified amino acids are expected to be poor substrates for proteases. The transport route of gluten derived peptides is still unknown, but hopefully, small peptide-like blockers would use the same route. The biggest challenge however remains: the generation of a peptide with such a high affinity that it can block in physiological concentrations selectively all of the HLA-DQ2.5 molecules in the lamina propria. This peptide is not likely to be composed of natural amino acids, but it must contain optimized non-natural amino acids in some if not all positions of the binding frame.

If we would succeed in designing an effective blocker for this very well described HLA molecule, it would not only be of benefit for CD patients, but it would also give us valuable information about a possible new treatment option for other HLA associated diseases.

References

- Babbitt BP, Allen PM, Matsueda G, Haber E, Unanue ER (1985) Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 317:359-361
- Barnes K, Leon F, Briand JP, Travers PJ, Hannestad K (1999) N-terminal elongation of a peptide determinant beyond the first primary anchor improves binding to H-2 I-Ad and HLA-DR1 by backbone-dependent and aromatic side chain-dependent interactions, respectively. *Eur. J. Immunol.* 29:189-195
- Ben-Horin S, Green PH, Bank I, Chess L, Goldstein I (2006) Characterizing the circulating, gliadin-specific CD4+ memory T cells in patients with celiac disease: linkage between memory function, gut homing and Th1 polarization. *J. Leukoc. Biol.* 79:676-685
- Bergseng E, Xia J, Kim CY, Khosla C, Sollid LM (2005) Main chain hydrogen bond interactions in the binding of proline-rich gluten peptides to the celiac disease associated HLA-DQ2 molecule. *J. Biol. Chem.* 280:21791-21796
- Bertoletti A, Sette A, Chisari FV, Penna A, Levrero M, De CM, Fiaccadori F, Ferrari C (1994) Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* 369:407-410
- Bhagat G, Naiyer AJ, Shah JG, Harper J, Jabri B, Wang TC, Green PH, Manavalan JS (2008) Small intestinal CD8+TCRgammadelta+NKG2A+ intraepithelial lymphocytes have attributes of regulatory cells in patients with celiac disease. *J. Clin. Invest* 118:281-293
- Bielekova B, Martin R (2001) Antigen-specific immunomodulation via altered peptide ligands. *J Mol Med* 79:552-565
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33-39
- Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO (1995) Receptor-ligand interactions measured by an improved spun column chromatography technique. A high efficiency and high throughput size separation method. *Biochim. Biophys. Acta* 1243:453-460
- Christinck ER, Luscher MA, Barber BH, Williams DB (1991) Peptide binding to class I MHC on living cells and quantitation of complexes required for CTL lysis. *Nature* 352:67-70
- Ciclitira PJ, Evans DJ, Fagg NL, Lennox ES, Dowling RH (1984) Clinical testing of gliadin fractions in coeliac patients. *Clin Sci (Lond)* 66:357-364
- Cooke WT, Smith WT (1966) Neurological disorders associated with adult coeliac disease. *Brain* 89:683-722
- Corrao G, Corazza GR, Bagnardi V, Brusco G, Ciacci C, Cottone M, Sategna GC, Usai P, Cesari P, Pelli MA, Loperfido S, Volta U, Calabro A, Certo M (2001) Mortality in patients with coeliac disease and their relatives: a cohort study. *Lancet* 358:356-361
- Cosnes J, Cellier C, Viola S, Colombel JF, Michaud L, Sarles J, Hugot JP, Ginies JL, Dabadie A, Mouterde O, Allez M, Nion-Larmurier I (2008) Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *Clin. Gastroenterol. Hepatol.* 6:753-758
- Crawford F, Kozono H, White J, Marrack P, Kappler J (1998) Detection of antigen-specific T cells with multivalent soluble class II MHC covalent peptide complexes. *Immunity.* 8:675-682

- Cresswell P (1994) Assembly, transport, and function of MHC class II molecules. *Annu. Rev. Immunol.* 12:259-293
- Cresswell P (1996) Invariant chain structure and MHC class II function. *Cell* 84:505-507
- de Haan EC, Moret EE, Wagenaar-Hilbers JPA, Liskamp RMJ, Wauben MHM (2005) Possibilities and limitations in the rational design of modified peptides for T cell mediated immunotherapy. *Mol. Immunol.* 42:365-373
- de Haan EC, Wauben MHM, Grosfeld-Stulemeyer MC, Kruijtz JAW, Liskamp RMJ, Moret EE (2002) Major histocompatibility complex class II binding characteristics of peptoid-peptide hybrids. *Bioorganic & Medicinal Chemistry* 10:1939-1945
- De Magistris MT, Alexander J, Coggeshall M, Altman A, Gaeta FCA, Grey HM, Sette A (1992) Antigen analog major histocompatibility complexes act as antagonists of the T-cell receptor. *Cell* 68:625-634
- De Wall SL, Painter C, Stone JD, Bandaranayake R, Wiley DC, Mitchison TJ, Stern LJ, Dedecker BS (2006) Noble metals strip peptides from class II MHC proteins. *Nat. Chem. Biol.* 2:197-201
- Demotz S, Grey HM, Sette A (1990) The minimal number of class II MHC-antigen complexes needed for T cell activation. *Science* 249:1028-1030
- Denzin LK, Cresswell P (1995) HLA-DM induces CLIP dissociation from MHC Class II $\alpha\beta$ dimers and facilitates peptide loading. *Cell* 82:155-165
- Dewar DH, Amato M, Ellis HJ, Pollock EL, Gonzalez-Cinca N, Wieser H, Ciclitira PJ (2006) The toxicity of high molecular weight glutenin subunits of wheat to patients with coeliac disease. *Eur. J. Gastroenterol. Hepatol.* 18:483-491
- Di Cagno R, De Angelis M, Auricchio S, Greco L, Clarke C, De Vincenzi M, Giovannini C, D'Archivio M, Landolfo F, Parrilli G, Minervini F, Arendt E, Gobbetti M (2004) Sourdough bread made from wheat and nontoxic flours and started with selected lactobacilli is tolerated in celiac sprue patients. *Applied and Environmental Microbiology* 70:1088-1096
- Dicke WK, Weijers HA, van de KAMER JH (1953) Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr.* 42:34-42
- Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Riecken EO, Schuppan D (1997) Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* 3:797-801
- Dieterich W, Laag E, Schopper H, Volta U, Ferguson A, Gillett H, Riecken EO, Schuppan D (1998) Autoantibodies to tissue transglutaminase as predictors of celiac disease. *Gastroenterology* 115:1317-1321
- Dube C, Rostom A, Sy R, Cranney A, Saloojee N, Garrity C, Sampson M, Zhang L, Yazdi F, Mamaladze V, Pan I, Macneil J, Mack D, Patel D, Moher D (2005) The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* 128:57-67
- Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zernakova A, Heap GA, Adany R, Aromaa A, Bardella MT, van den Berg LH, Bockett NA, de la Concha EG, Dema B, Fehrmann RS, Fernandez-Arquero M, Fiatal S, Grandone E, Green PM, Groen HJ, Gwilliam R, Houwen RH, Hunt SE, Kaukinen K, Kelleher D, Korponay-Szabo I, Kurppa K, Macmathuna P, Maki M, Mazzilli MC, McCann OT, Mearin ML, Mein CA, Mirza MM, Mistry V, Mora B, Morley KI, Mulder CJ, Murray JA, Nunez C, Oosterom E, Ophoff RA, Polanco I, Peltonen L, Platteel M, Rybak A, Salomaa V, Schweizer JJ, Sperandeo MP, Tack GJ, Turner G, Veldink JH, Verbeek WH, Weersma RK, Wolters VM, Urcelay E, Cukrowska B, Greco L, Neuhausen SL, McManus R, Barisani D, Deloukas P, Barrett JC, Saavalainen P, Wijmenga C, van Heel DA (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* 42:295-302

- Ettinger RA, Papadopoulos GK, Moustakas AK, Nepom GT, Kwok WW (2006) Allelic variation in key peptide-binding pockets discriminates between closely related diabetes-protective and diabetes-susceptible HLA-DQB1*06 alleles. *J Immunol* 176:1988-1998
- Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1994) Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics* 39:230-242
- Fallang LE, Bergseng E, Hotta K, Berg-Larsen A, Kim CY, Sollid LM (2009) Differences in the risk of celiac disease associated with HLA-DQ2.5 or HLA-DQ2.2 are related to sustained gluten antigen presentation. *Nat. Immunol.* 10:1096-1101
- Fasano A, Catassi C (2001) Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum. *Gastroenterology* 120:636-651
- Ferguson A, Murray D (1971) Quantitation of intraepithelial lymphocytes in human jejunum. *Gut* 12:988-994
- Fleckenstein B, Molberg Ø, Qiao SW, Schmid DG, von der Mülbe F, Elgstøen K, Jung G, Sollid LM (2002) Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. *J. Biol. Chem.* 277:34109-34116
- Frisoni M, Corazza GR, Lafiandra D, De Ambrogio E., Filipponi C, Bonvicini F, Borasio E, Porceddu E, Gasbarrini G (1995) Wheat deficient in gliadins: promising tool for treatment of coeliac disease. *Gut* 36:375-378
- Gass J, Bethune MT, Siegel M, Spencer A, Khosla C (2007) Combination enzyme therapy for gastric digestion of dietary gluten in patients with celiac sprue. *Gastroenterology* 133:472-480
- Gass J, Khosla C (2007) Prolyl endopeptidases. *Cell Mol. Life Sci* 64:345-355
- Ghosh P, Amaya M, Mellins E, Wiley DC (1995) The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. *Nature* 378:457-462
- Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, Fulchignoni-Lataud MC, Zavattari P, Momigliano-Richiardi P, Casari G, Gasparini P, Tosi R, Mantovani V, De VS, Iacono G, D'Alfonso A, Selinger-Leneman H, Lemaître A, Serre JL, Clerget-Darpoux F (2001) Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. *Ann. Hum Genet.* 65:35-41
- Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R, D'Agate C, Tinto N, Sacchetti L, Tosi R, Stazi MA (2002) The first large population based twin study of coeliac disease. *Gut* 50:624-628
- Greenwood FC, Hunter WM, Glover JS (1963) The preparation of I-131-labelled human growth hormone of high specific radioactivity. *Biochem. J* 89:114-123
- Guy-Grand D, DiSanto JP, Henchoz P, Malassis-Seris M, Vassalli P (1998) Small bowel enteropathy: role of intraepithelial lymphocytes and of cytokines (IL-12, IFN-gamma, TNF) in the induction of epithelial cell death and renewal. *Eur. J. Immunol.* 28:730-744
- Hammer J, Takacs B, Sinigaglia F (1992) Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J. Exp. Med.* 176:1007-1013
- Harding CV, Unanue ER (1990) Quantitation of antigen-presenting cell MHC class II/peptide complexes necessary for T-cell stimulation. *Nature* 346:574-576
- Hausch F, Shan L, Santiago NA, Gray GM, Khosla C (2002) Intestinal digestive resistance of immunodominant gliadin peptides. *Am. J. Physiol Gastrointest. Liver Physiol* 283:996-1003

- Högberg L, Grodzinsky E, Stenhammar L (2003) Better dietary compliance in patients with coeliac disease diagnosed in early childhood. *Scand. J Gastroenterol.* 38:751-754
- Holmes GK, Prior P, Lane MR, Pope D, Allan RN (1989) Malignancy in coeliac disease--effect of a gluten free diet. *Gut* 30:333-338
- Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC, Jr., Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S (2004) Gene map of the extended human MHC. *Nat Rev. Genet.* 5:889-899
- Hüe S, Mention JJ, Monteiro RC, Zhang SL, Cellier C, Schmitz J, Verkarre V, Fodil N, Bahram S, Cerf-Bensussan N, Caillat-Zucman S (2004) A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity* 21:367-377
- Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA (2008) Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet.* 40:395-402
- Jabri B, de Serre NP, Cellier C, Evans K, Gache C, Carvalho C, Mougenot JF, Allez M, Jian R, Desreumaux P, Colombel JF, Matuchansky C, Cugnenc H, Lopez-Botet M, Vivier E, Moretta A, Roberts AI, Ebert EC, Guy-Grand D, Brousse N, Schmitz J, Cerf-Bensussan N (2000) Selective expansion of intraepithelial lymphocytes expressing the HLA-E-specific natural killer receptor CD94 in celiac disease. *Gastroenterology* 118:867-879
- Jabri B, Sollid LM (2006) Mechanisms of disease: immunopathogenesis of celiac disease. *Nat. Clin. Pract. Gastroenterol. Hepatol.* 3:516-525
- Jabri B, Sollid LM (2009) Tissue-mediated control of immunopathology in coeliac disease. *Nat. Rev. Immunol.* 9:858-870
- Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1996) Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc. Natl. Acad. Sci. U. S. A* 93:734-738
- Jardetzky TS, Gorga JC, Busch R, Rothbard J, Strominger JL, Wiley DC (1990) Peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. *EMBO J.* 9:1797-1803
- Johansen BH, Buus S, Vartdal F, Viken H, Eriksen JA, Thorsby E, Sollid LM (1994) Binding of peptides to HLA-DQ molecules: peptide binding properties of the disease-associated HLA-DQ(α 1*0501, β 1*0201) molecule. *Int. Immunol.* 6:453-461
- Johansen BH, Gjertsen HA, Vartdal F, Buus S, Thorsby E, Lundin KE, Sollid LM (1996a) Binding of peptides from the N-terminal region of α -gliadin to the celiac disease-associated HLA-DQ2 molecule assessed in biochemical and T cell assays. *Clin. Immunol. Immunopathol.* 79:288-293
- Johansen BH, Jensen T, Thorpe CJ, Vartdal F, Thorsby E, Sollid LM (1996b) Both α and β chain polymorphisms determine the specificity of the disease-associated HLA-DQ2 molecules, with beta chain residues being most influential. *Immunogenetics* 45:142-150
- Jüse U, Fleckenstein B, Bergseng E, Sollid LM (2009) Soluble HLA-DQ2 expressed in S2 cells copurifies with a high affinity insect cell derived protein. *Immunogenetics* 61:81-89

Kapoerchan VV, Wiesner M, Hillaert U, Drijfhout JW, Overhand M, Alard P, van der Marel GA, Overkleef HS, Koning F (2010) Design, synthesis and evaluation of high-affinity binders for the celiac disease associated HLA-DQ2 molecule. *Mol. Immunol.* 47:1091-1097

Kapoerchan VV, Wiesner M, Overhand M, van der Marel GA, Koning F, Overkleef HS (2008) Design of azidoproline containing gluten peptides to suppress CD4(+) T-cell responses associated with Celiac disease. *Bioorg. Med. Chem.* 16:2053-2062

Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, Ciclitira PJ, Sollid LM, Partanen J (2003) HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the european genetics cluster on celiac disease. *Hum. Immunol.* 64:469-477

Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM (2004) Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc. Natl. Acad. Sci. U. S. A* 101:4175-4179

Kozono H, White J, Clements J, Marrack P, Kappler J (1994) Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* 369:151-154

Krebs S, Folkers G, Rognan D (1998) Binding of rationally designed non-natural peptides to the human leukocyte antigen HLA-B*2705. *J. Pept. Sci.* 4:378-388

Krieger JI, Karr RW, Grey HM, Yu WY, O'Sullivan D, Batovsky L, Zheng ZL, Colon SM, Gaeta FC, Sidney J, . (1991) Single amino acid changes in DR and antigen define residues critical for peptide-MHC binding and T cell recognition. *J. Immunol.* 146:2331-2340

Lam KS, Lehman AL, Song A, Doan N, Enstrom AM, Maxwell J, Liu R (2003) Synthesis and screening of "one-bead one-compound" combinatorial peptide libraries. *Methods Enzymol.* 369:298-322

Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* 354:82-84

Leffler DA, Edwards-George J, Dennis M, Schuppan D, Cook F, Franko DL, Blom-Hoffman J, Kelly CP (2008) Factors that influence adherence to a gluten-free diet in adults with celiac disease. *Dig. Dis. Sci* 53:1573-1581

Lerner A, Kumar V, Iancu TC (1994) Immunological diagnosis of childhood coeliac disease: comparison between antigliadin, antireticulin and antiendomysial antibodies. *Clin. Exp. Immunol.* 95:78-82

Lundin KE, Scott H, Fausa O, Thorsby E, Sollid LM (1994) T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. *Hum. Immunol.* 41:285-291

Lundin KE, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, Thorsby E, Sollid LM (1993) Gliadin-specific, HLA-DQ(α 1*0501, β 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *J. Exp. Med.* 178:187-196

Madden DR (1995) The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol* 13:587-622

Maiuri L, Ciacci C, Auricchio S, Brown V, Quarantino S, Londei M (2000) Interleukin 15 mediates epithelial changes in celiac disease. *Gastroenterology* 119:996-1006

Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raia V, Auricchio S, Picard J, Osman M, Quarantino S, Londei M (2003) Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet* 362:30-37

- Maki M, Collin P (1997) Coeliac disease. *Lancet* 349:1755-1759
- Maki M, Mustalahti K, Kokkonen J, Kulmala P, Haapalahti M, Karttunen T, Ilonen J, Laurila K, Dahlbom I, Hansson T, Hopfl P, Knip M (2003) Prevalence of celiac disease among children in Finland. *N. Engl. J. Med.* 348:2517-2524
- Marsh MN, Swift JA, Williams ED (1968) Studies of small-intestinal mucosa with the scanning electron microscope. *Br. Med J* 4:95-96
- McFarland BJ, Beeson C, Sant AJ (1999) Cutting edge: A single, essential hydrogen bond controls the stability of peptide-MHC class II complexes. *Journal of Immunology* 163:3567-3571
- Merese B, Curran SA, Ciszewski C, Orbelyan G, Setty M, Bhagat G, Lee L, Tretiakova M, Semrad C, Kistner E, Winchester RJ, Braud V, Lanier LL, Geraghty DE, Green PH, Guandalini S, Jabri B (2006) Reprogramming of CTLs into natural killer-like cells in celiac disease. *J. Exp. Med.* 203:1343-1355
- Merese B, Ripoché J, Heyman M, Cerf-Bensussan N (2009) Celiac disease: from oral tolerance to intestinal inflammation, autoimmunity and lymphomagenesis. *Mucosal. Immunol.* 2:8-23
- Molberg Ø, Kett K, Scott H, Thorsby E, Sollid LM, Lundin KE (1997) Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. *Scand. J Immunol* 46:103-108
- Molberg Ø, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, Fugger L, Scott H, Noren O, Roepstorff P, Lundin KE, Sjöström H, Sollid LM (1998) Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* 4:713-717
- Molberg Ø, Uhlen AK, Jensen T, Flaete NS, Fleckenstein B, Arentz-Hansen H, Raki M, Lundin KE, Sollid LM (2005) Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. *Gastroenterology* 128:393-401
- Molteni N, Caraceni MP, Bardella MT, Ortolani S, Gandolini GG, Bianchi P (1990) Bone mineral density in adult celiac patients and the effect of gluten-free diet from childhood. *Am. J. Gastroenterol.* 85:51-53
- Nelson CA, Fremont DH (1999) Structural principles of MHC class II antigen presentation. *Rev. Immunogenet.* 1:47-59
- Nilsen EM, Jahnsen FL, Lundin KE, Johansen FE, Fausa O, Sollid LM, Jahnsen J, Scott H, Brandtzaeg P (1998) Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology* 115:551-563
- Nilsen EM, Lundin KE, Krajci P, Scott H, Sollid LM, Brandtzaeg P (1995) Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* 37:766-776
- Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, Paparo F, D'Alfonso S, Giordano M, Sferlazzas C, Magazzu G, Momigliano-Richiardi P, Greco L, Stazi MA (2006) Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut* 55:803-808
- O'Sullivan D, Sidney J, Appella E, Walker L, Phillips L, Colon SM, Miles C, Chesnut RW, Sette A (1990) Characterization of the specificity of peptide binding to four DR haplotypes. *J Immunol* 145:1799-1808
- Oberhuber G, Granditsch G, Vogelsang H (1999) The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur. J. Gastroenterol. Hepatol.* 11:1185-1194

Parrot I, Huang PC, Khosla C (2002) Circular dichroism and nuclear magnetic resonance spectroscopic analysis of immunogenic gluten peptides and their analogs. *J. Biol. Chem.* 277:45572-45578

Paterson BM, Lammers KM, Arrieta MC, Fasano A, Meddings JB (2007) The safety, tolerance, pharmacokinetic and pharmacodynamic effects of single doses of AT-1001 in coeliac disease subjects: a proof of concept study. *Aliment. Pharmacol. Ther.* 26:757-766

Perera L, Shao L, Patel A, Evans K, Meresse B, Blumberg R, Geraghty D, Groh V, Spies T, Jabri B, Mayer L (2007) Expression of nonclassical class I molecules by intestinal epithelial cells. *Inflamm. Bowel. Dis.* 13:298-307

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567

Pinier M, Verdu EF, Nasser-Eddine M, David CS, Vezina A, Rivard N, Leroux JC (2009) Polymeric binders suppress gliadin-induced toxicity in the intestinal epithelium. *Gastroenterology* 136:288-298

Price P, Witt C, Allcock R, Sayer D, Garlepp M, Kok CC, French M, Mallal S, Christiansen F (1999) The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol Rev.* 167:257-274

Qiao SW, Bergseng E, Molberg Ø, Jung G, Fleckenstein B, Sollid LM (2005) Refining the rules of gliadin T cell epitope binding to the disease associated DQ2 molecule in celiac disease; importance of proline spacing and glutamine deamidation. *J. Immunol.* 175:254-261

Qiao SW, Sollid LM, Blumberg RS (2009) Antigen presentation in celiac disease. *Curr. Opin. Immunol.* 21:111-117

Quarsten H, McAdam SN, Jensen T, Arentz-Hansen H, Molberg Ø, Lundin KE, Sollid LM (2001) Staining of celiac disease-relevant T cells by peptide-DQ2 multimers. *J. Immunol.* 167:4861-4868

Quarsten H, Paulsen G, Johansen BH, Thorpe CJ, Holm A, Buus S, Sollid LM (1998) The P9 pocket of HLA-DQ2 (non-Aspβ57) has no particular preference for negatively charged anchor residues found in other type 1 diabetes-predisposing non-Aspβ57 MHC class II molecules. *Int. Immunol.* 10:1229-1236

Rammensee HG, Friede T, Stevanovic S (1995) MHC ligands and peptide motifs: first listing. *Immunogenetics* 41:178-228

Reddy A, Fried B (2009) An update on the use of helminths to treat Crohn's and other autoimmune diseases. *Parasitol. Res.* 104:217-221

Roberts AI, Lee L, Schwarz E, Groh V, Spies T, Ebert EC, Jabri B (2001) NKG2D receptors induced by IL-15 costimulate CD28-negative effector CTL in the tissue microenvironment. *J Immunol* 167:5527-5530

Roche PA, Marks MS, Cresswell P (1991) Formation of a nine-subunit complex by HLA class II glycoproteins and the invariant chain. *Nature* 354:392-394

Sette A, Sidney J, Albertson M, Miles C, Colon SM, Pedrazzini T, Lamont AG, Grey HM (1990) A novel approach to the generation of high affinity class II-binding peptides. *J. Immunol.* 145:1809-1813

Shan L, Molberg Ø, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM, Khosla C (2002) Structural basis for gluten intolerance in celiac sprue. *Science* 297:2275-2279

Sher KS, Mayberry JF (1994) Female fertility, obstetric and gynaecological history in coeliac disease. A case control study. *Digestion* 55:243-246

- Siegel M, Bethune MT, Gass J, Ehren J, Xia J, Johannsen A, Stuge TB, Gray GM, Lee PP, Khosla C (2006) Rational design of combination enzyme therapy for celiac sprue. *Chem. Biol.* 13:649-658
- Siegel M, Khosla C (2007) Transglutaminase 2 inhibitors and their therapeutic role in disease states. *Pharmacol. Ther.* 115:232-245
- Siegel M, Xia J, Khosla C (2007) Structure-based design of α -amido aldehyde containing gluten peptide analogues as modulators of HLA-DQ2 and transglutaminase 2. *Bioorg. Med Chem.* 15:6253-6261
- Sloan-Lancaster J, Allen PM (1996) Altered peptide ligand-induced partial T cell activation: molecular mechanisms and role in T cell biology. *Annu. Rev. Immunol.* 14:1-27
- Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW (1998) Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. *J Exp Med* 188:1511-1520
- Sollid LM (2002) Coeliac disease: dissecting a complex inflammatory disorder. *Nat. Rev. Immunol.* 2:647-655
- Sollid LM, Lie BA (2005) Celiac disease genetics: current concepts and practical applications. *Clin. Gastroenterol. Hepatol.* 3:843-851
- Sollid LM, Lundin KE (2009) Diagnosis and treatment of celiac disease. *Mucosal. Immunol.* 2:3-7
- Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J. Exp. Med.* 169:345-350
- Sollid LM, Molberg Ø, McAdam S, Lundin KE (1997) Autoantibodies in coeliac disease: tissue transglutaminase--guilt by association? *Gut* 41:851-852
- Sollid LM, Thorsby E (1993) HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 105:910-922
- Spaenij-Dekking L, Kooy-Winkelaar Y, Koning F (2005a) The Ethiopian cereal tef in celiac disease. *N. Engl. J Med* 353:1748-1749
- Spaenij-Dekking L, Kooy-Winkelaar Y, van VP, Drijfhout JW, Jonker H, van SL, Smulders MJ, Bosch D, Gilissen LJ, Koning F (2005b) Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. *Gastroenterology* 129:797-806
- Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E (1992) HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. *Hum. Immunol.* 35:188-192
- Stepniak D, Spaenij-Dekking L, Mitea C, Moester M, de RA, Baak-Pablo R, van VP, Edens L, Koning F (2006) Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease. *Am. J. Physiol Gastrointest. Liver Physiol* 291:621-629
- Stepniak D, Wiesner M, de Ru AH, Moustakas AK, Drijfhout JW, Papadopoulos GK, Van Veelen PA, Koning F (2008) Large-Scale Characterization of Natural Ligands Explains the Unique Gluten-Binding Properties of HLA-DQ2. *J Immunol* 180:3268-3278
- Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal-structure of the human class-II MHC protein HLA-DR1 complexed with an influenza-virus peptide. *Nature* 368:215-221
- Stern LJ, Wiley DC (1992) The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. *Cell* 68:465-477

- Stevanovic S, Schild H (1999) Quantitative aspects of T cell activation--peptide generation and editing by MHC class I molecules. *Semin. Immunol.* 11:375-384
- Sykulev Y, Joo M, Vturina I, Tsomides TJ, Eisen HN (1996) Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity.* 4:565-571
- Thorsby E (1997) Invited anniversary review: HLA associated diseases. *Hum. Immunol.* 53:1-11
- Thorsby E, Lie BA (2005) HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. *Transpl. Immunol.* 14:175-182
- Todd JA, Bell JI, McDevitt HO (1987) HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 329:599-604
- Trier JS (1991) Celiac sprue. *N. Engl. J. Med.* 325:1709-1719
- Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, Benckhuijsen W, Pena S, Mearin L, Drijfhout JW, Koning F (2002) The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122:1729-1737
- van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, Houwen RH, Wijmenga C (2003) A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 125:1032-1041
- van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Koning F (1996) Peptide binding characteristics of the coeliac disease-associated DQ(α 1*0501, β 1*0201) molecule. *Immunogenetics* 44:246-253
- van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Papadopoulos GK, Koning F (1997) Unique peptide binding characteristics of the disease-associated DQ(α 1*0501, β 1*0201) vs the non-disease-associated DQ(α 1*0201, β 1*0202) molecule. *Immunogenetics* 46:484-492
- van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C (2007) A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet.* 39:827-829
- van Heel DA, Hunt K, Greco L, Wijmenga C (2005) Genetics in coeliac disease. *Best. Pract. Res. Clin. Gastroenterol.* 19:323-339
- Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanovic S, Eriksen JE, Sletten K, Thorsby E, Rammensee HG, Sollid LM (1996) The peptide binding motif of the disease associated HLA-DQ (α 1*0501, β 1*0201) molecule. *Eur. J. Immunol.* 26:2764-2772
- Ventura A, Magazzu G, Greco L (1999) Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. *Gastroenterology* 117:297-303
- Viken HD, Paulsen G, Sollid LM, Lundin KEA, Tjonnfjord GE, Thorsby E, Gaudernack G (1995) Characterization of an HLA-DQ2-specific monoclonal antibody - Influence of amino acid substitutions in DQ β 1*0202. *Hum Immunol* 42:319-327
- Wallny HJ, Sollami G, Karjalainen K (1995) Soluble mouse major histocompatibility complex class II molecules produced in Drosophila cells. *Eur. J. Immunol.* 25:1262-1266
- Wettstein DA, Boniface JJ, Reay PA, Schild H, Davis MM (1991) Expression of a class II major histocompatibility complex (MHC) heterodimer in a lipid-linked form with enhanced peptide/soluble MHC complex formation at low pH. *J. Exp. Med.* 174:219-228

Wieser H (1995) The precipitating factor in coeliac disease. *Baillieres Clin. Gastroenterol.* 9:191-207

Wieser H, Mödl A, Seilmeier W, Belitz HD (1987) High-performance liquid chromatography of gliadins from different wheat varieties: Amino acid composition and N-terminal amino acid sequence of components. *Z. Lebensm. Unters. Forsch.* 185:371-378

Wiesner M, Stepniak D, de Ru AH, Moustakis AK, Drijfhout JW, Papadopoulos GK, Van Veelen PA, Koning F (2008) Dominance of an alternative CLIP sequence in the celiac disease associated HLA-DQ2 molecule. *Immunogenetics* 60:551-555

Xia J, Bergseng E, Fleckenstein B, Siegel M, Kim CY, Khosla C, Sollid LM (2007) Cyclic and dimeric gluten peptide analogues inhibiting DQ2-mediated antigen presentation in celiac disease. *Bioorg. Med. Chem.* 15:6565-6573

Zelnik N, Pacht A, Obeid R, Lerner A (2004) Range of neurologic disorders in patients with celiac disease. *Pediatrics* 113:1672-1676

Errata

Paper I:

Table 1: the correct sequence for peptide P198 is KPLLIIAEDVEGEY

Paper I

**Soluble HLA-DQ2 expressed in S2 cells copurifies with a high affinity insect cell
derived protein**

Ulrike Jüse, Burkhard Fleckenstein, Elin Bergseng, Ludvig M. Sollid
Immunogenetics 2009 Feb;61(2):81-9

Soluble HLA-DQ2 expressed in S2 cells copurifies with a high affinity insect cell derived protein

Ulrike Jüse · Burkhard Fleckenstein · Elin Bergseng ·
Ludvig M. Sollid

Received: 18 July 2008 / Accepted: 14 October 2008 / Published online: 6 November 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract We here describe that soluble HLA-DQ2 (sDQ2) molecules, when expressed in *Drosophila melanogaster* S2 insect cells without a covalently tethered peptide, associate tightly with the *D. melanogaster* calcium binding protein DCB-45. The interaction between the proteins is stable in S2 cell culture and during affinity purification, which is done at high salt concentrations and pH 11.5. After affinity purification, the sDQ2/DCB-45 complex exists in substantial quantities next to a small amount of free heterodimeric sDQ2 and large amounts of aggregated sDQ2 free of DCB-45. Motivated by the stable complex formation and our interest in the development of reagents which inhibit HLA-DQ2 peptide binding, we have further characterized the sDQ2/DCB-45 interaction. Several lines of evidence indicate that an N-terminal fragment of DCB-45 is involved in the interaction with the peptide binding groove of sDQ2. Further mapping of this fragment of 54 residues identified a pentadecapeptide with high affinity for sDQ2 which may serve as a lead compound for the design of HLA-DQ2 blockers.

Keywords HLA · sDQ2 · DCB-45 · S2 ·
Protein interactions · Peptide binding

Introduction

HLA-DQ2 (DQA1*0501/DQB1*0201) is associated with several immune-mediated disorders including celiac disease (Sollid et al. 1989; Thorsby and Ronningen 1993; Todd et al. 1987). In order to develop specific blockers as a potential future treatment for celiac disease and other HLA-DQ2 associated diseases, there is a need to define high affinity ligands for HLA-DQ2. The development of such blockers will be facilitated by the production of recombinant HLA-DQ2 that can be made in high quantities and utilized for binding studies. On this background, we have produced recombinant, soluble HLA-DQ2 (sDQ2) molecules in either stably transfected S2 cells or in baculovirus infected Sf9 cells. Interestingly, when expressing sDQ2 in S2 cells without a covalently bound peptide that occupies the binding groove, we obtained remarkably high amounts of stable molecules. Further analysis of these sDQ2 molecules revealed that a major proportion of the molecules copurifies with the *Drosophila melanogaster* calcium binding protein DCB-45, a homolog to the *D. melanogaster* supercoiling factor (SCF; Kobayashi et al. 1998). Given the apparent ability of the protein to stabilize sDQ2, we hypothesized that it could contain a peptide which binds efficiently to the binding groove of sDQ2 and which could function as a lead peptide for the development of HLA-DQ2 blockers. Here, we report on the identification and characterization of the DCB-45 sequence which mediates binding to sDQ2. Next to characterizing the interaction between sDQ2 and DCB-45, our observations are of general relevance for expression of recombinant major histocompatibility complex (MHC) class II molecules in S2 cells.

U. Jüse (✉) · B. Fleckenstein · E. Bergseng · L. M. Sollid
Centre for Immune Regulation, Institute of Immunology,
University of Oslo,
Oslo, Norway
e-mail: ulrike.juse@medisin.uio.no

U. Jüse · L. M. Sollid
Institute of Immunology, Rikshospitalet University Hospital,
0027 Oslo, Norway

Materials and methods

HLA expression and purification

Water soluble HLA-DQ2 (DQA1*0501/DQB1*0201) was expressed in a *D. melanogaster* S2 cell line by cotransfection of three vectors: a pMtal vector with the sequence encoding for the extracellular part of the DQ α -chain fused to the Fos zipper, a pMtal vector with the sequence encoding the extracellular part of the DQ β -chain fused to the Jun zipper, and a pCoHYGRO resistance vector (Paulsen et al., unpublished). The construct did not include a sequence for a high affinity peptide ligand tethered to the β -chain. The stably transfected S2 cells were grown in 1L cell spin flasks at 22°C in serum-free media (Insect-XPRESS™, BioWhittaker, Walkersville, MD, USA), containing 300 μ g/ml hygromycin B (Invitrogen, Carlsbad, CA, USA) and 25 μ g/ml garamycin (Schering-Plough, Kenilworth, NJ, USA). The pMtal vectors contain a metallothionein promoter and the production of sDQ2 was induced by 100 mM CuSO₄ over 3 days. From cell culture supernatants, the sDQ2 molecules were affinity-purified like previously described (Quarsten et al. 2001). HLA-DM molecules were produced as soluble molecules in transfected S2 cells (kind gift of Elizabeth Mellins) and purified by FLAG-tag immunoaffinity chromatography and size exclusion chromatography as described (Sloan et al. 1995). Detergent-solubilized HLA-DQ2 (DQA1*0501/DQB1*0201; EBV-DQ2) molecules were purified from Epstein–Barr virus-transformed B lymphoblastoid cell lines as previously described (Johansen et al. 1994). The protein concentration of the various HLA molecules was determined by a BCA protein assay kit (Pierce, Rockford, IL, USA).

Size exclusion chromatography and gel electrophoresis

Preparative size exclusion chromatography was performed on an Äkta purifier system (Amersham Biosciences Corp., Piscataway, NJ, USA) using a Superdex 200 10/300 GL column (Amersham Bioscience). Proteins were separated by isocratic elution (flow rate 0.75 ml/min) using phosphate buffered saline (PBS; pH 7.3) and monitored at 280 nm. One-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using 6–12% polyacrylamide gels. Samples were reduced by heating in Laemmli buffer containing 1% β -mercaptoethanol (95°C, 5 min). Proteins were stained by Coomassie Blue.

Trypsin treatment

For protein identification by peptide mass fingerprinting, protein bands were excised from the Coomassie Blue stained gel and in-gel digested as previously described

(Fleckenstein et al. 2004). For the acetylation of primary amino groups, the gel pieces were incubated in 3% acetanhydride/0.1 M NaHCO₃ pH 8.4 for 2 h at room temperature, followed by several washing steps and tryptic digestion. Partial digestion of the sDQ2/DCB-45 complex in solution was performed with *N*-tosyl L-phenylalanyl chloromethyl ketone (TPCK) trypsin agarose beads (Pierce). For 24 μ g protein, 4.8 μ l bead suspension in 0.1 M NH₄HCO₃ buffer in a total volume of 22.8 μ l was used. The samples were incubated for 17 min at 37°C under rotation. For size exclusion chromatography, the beads were removed by filtration.

Mass spectrometry

Matrix-assisted laser desorption–ionization time-of-flight mass spectra (MALDI-TOF MS) were acquired on a MALDI-TOF/TOF instrument (Ultraflex II, Bruker Daltonics, Bremen, Germany). Tryptic peptide mixtures and peptides eluted from sDQ2 were desalted and concentrated on Poros 20 R2 reverse-phase packing sorbent (Applied Biosystems, Foster City, CA, USA) packed in 20- μ l GELoader tips (Eppendorf, Hamburg, Germany). Peptides were eluted onto a stainless steel target plate using 70% acetonitrile, containing 0.1% trifluoroacetic acid (TFA) and 10 g/l α -cyano-4-hydroxycinnamic acid. Synthetic peptides were also analyzed using α -cyano-4-hydroxycinnamic acid, but then samples were applied as a dried droplet.

Peptide synthesis

Peptides with a length of 11–20 amino acid residues were synthesized on Rink amid methylbenzhydrylamine-resin using Fmoc/2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) chemistry and a pipetting robot (Syrro I, MultiSynTech, Bochum, Germany). Coupling was performed with a tenfold molar excess of each Fmoc-L-amino acid and HBTU and a 15-fold excess of *N,N*-diisopropylethylamine. Piperidine (20% in dimethylformamide) was used for Fmoc deprotection and 96% TFA containing 2% triisopropylsilane and 2% water were used for side chain deprotection and cleavage from the resin. The identity was confirmed by MALDI-TOF MS, and the purity was determined by analytical reversed-phase high performance liquid chromatography (HPLC; Agilent 1100 system, Agilent Technologies, Santa Clara, CA, USA) using a Zorbax C18 column (Agilent Technologies). The 54 residue long peptide was synthesized on 10 mg NovaPEG Rink Amide resin LL (Novabiochem®, Merck KGaA, Darmstadt, Germany) with a low loading capacity of 0.16 mmol/g. The pseudoproline dipeptides Fmoc-Gly-(Hmb)Gly-OH (Novabiochem) in position 24 and Fmoc-Glu(OtBu)-Ser($\psi^{Me,Me}pro$)-OH (Novabiochem) in position

41 and 52 were used to substitute for Gly-Gly and Glu-Ser, respectively. The secondary amide bond of these amino acid surrogates is reversibly protected and cleaved during deprotection with 96% TFA. The first 30 cycles were run under standard conditions as described above. From cycle 30 on, all coupling steps were repeated and 30% piperidine in dimethylformamide for Fmoc deprotection was used. Side chain deprotection and cleavage from the resin was performed as described above. From the crude peptide, the observed TFA ester on the N-terminal Ser side chain was removed by treatment with 1 M NaOH for 30 min, followed by neutralization with equimolar amounts of HCl. The identity and purity were determined like described above.

Labeling of synthetic peptides

The HLA-DQ2 high affinity ligand P198 (KPLLIIAEDVE-GEY, *Mycobacterium bovis* 65 kDa Hsp 243–255Y) was either fluorescently labeled or radiolabeled. The radio-labeling with ^{125}I was done with the chloramine T method (Greenwood et al. 1963). Fluorescence labeling was done by N-terminally coupling of 5(6)-carboxyfluorescein (CF) using 5 equivalents CF and 6 equivalents *N,N*-diisopropylcarbodiimide. The coupling step was repeated several times until a negative Kaiser test was obtained (Kaiser et al. 1970). The fluorescently labeled peptide was cleaved from the resin and analyzed as described above.

Peptide binding assays

Detergent-solubilized HLA-DQ2 (DQA1*0501/DQB1*0201; EBV-DQ2) at concentrations of 0.1–0.2 μM was incubated with the radioactively labeled P198 indicator peptide and synthetic peptides at various concentrations as described earlier (Johansen et al. 1994). The EBV-DQ2-peptide complexes were subsequently separated from unbound peptides by size exclusion chromatography in a spin column system as described (Buus et al. 1995). The radioactivity in the void volume and in the column material was counted by a γ -counter (Wallac, Turku, Finland) and IC50 values were calculated. sDQ2 (6.25 μM) or trypsin digested sDQ2 (6.25 μM) was incubated with CF-labeled P198 indicator peptide (10 μM). In some instances, soluble HLA-DM (sDM) was added in a concentration of 0.38 μM . The incubation was performed in a total volume of 13 μl of a citrate phosphate buffer (pH 5.3) at 37°C over 48 h. Next, 5 μl of each sample were transferred on a Zorbax GF 450 4.9×250-mm column and eluted with 150 mM PBS at a flow rate of 1 ml/min using an Agilent 1100 HPLC system. The UV signal (214 nm) and the fluorescence signal (Ex. 490 nm, Em. 520 nm) were monitored. Peptide binding to sDQ2 was quantified by measuring the area under the curve (AUC) of

the fluorescence signal for peaks with a maximum between 9.5 and 9.7 min (corresponding to sDQ2 with bound indicator peptide). The mean values from at least two independent experiments were plotted.

Results

Characterization of the sDQ2 preparation by SDS-PAGE and size exclusion chromatography

The affinity purified sDQ2 was analyzed by SDS-PAGE under reducing conditions. In addition to the bands of 35 and 25 kDa corresponding to the sDQ2 α - and β -chains, respectively, a protein with an approximate molecular weight of 50 kDa was observed (Fig. 1A). In size exclusion chromatography under nonreducing conditions, the same sample eluted in three peaks, which were fractionated for further analysis by SDS-PAGE (Fig. 1B,C). As the first peak eluted with the void volume, we consider it to represent aggregates of sDQ2 with a size larger than 600 kDa. SDS-PAGE analysis of this fraction (Fig. 1C, lanes a and e) showed the presence of sDQ2 α - and β -chains, but no band at 50 kDa. In the second fraction (Fig. 1C, lanes b and f), the 50-kDa protein was detected with similar intensity as the sDQ2 α - and β -chains. Peak 3 mainly contains sDQ2 α - and β -chain (Fig. 1C, lane c). The small amount of the 50-kDa protein can be explained by the overlap of peak 2 and 3 and the small amounts of peak 2 present in fraction 3. The amount of protein applied to each lane of the SDS-PAGE was calculated by combining the known amount of total protein subjected to size exclusion chromatography and by determination of the peak areas for each of the fractions 1–3. This was 2.2 μg for each of the lanes a–c. For the lanes e and f, 29 μg of protein of the fractions 1 and 2, respectively, was loaded in each lane. The protein bands in the lanes a and e (corresponding to fraction 1) appeared weaker compared to the protein bands detected in the lanes b and f (corresponding to fraction 2), although identical amount of protein should have been loaded. This might be due to aggregation of sDQ2 in fraction 1, leading to an incomplete transfer into the gel.

Identification of the 50-kDa protein by peptide mass fingerprinting

In order to identify the nature of the 50-kDa contamination, the protein was subjected to tryptic in-gel digestion. The digest was analyzed on a MALDI-TOF/TOF mass spectrometer and a MASCOT search reported the identification of a *D. melanogaster*-derived SCF (accession number Q9W0H8_DROME). Acetylation prior to digestion and analysis by MALDI-TOF/TOF MS identified the tryptic

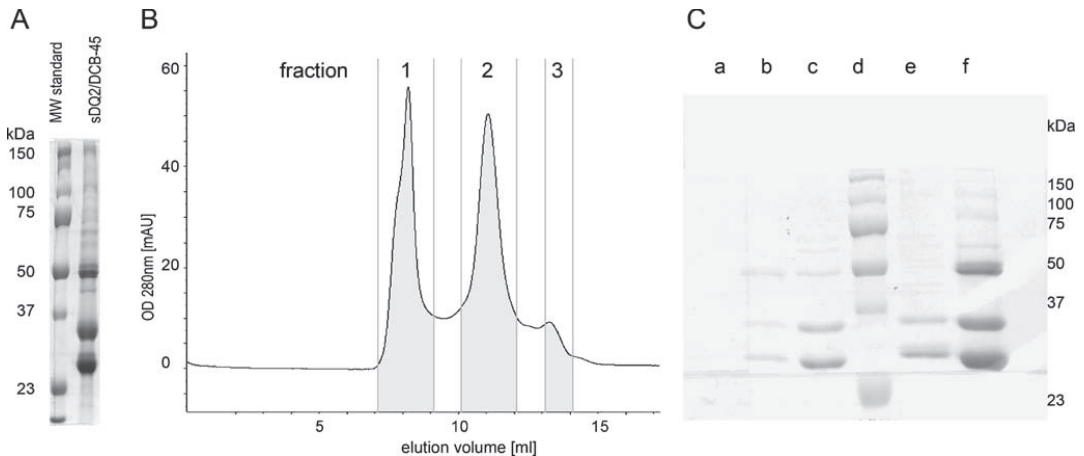


Fig. 1 Analysis of sDQ2 produced in S2 insect cells by reducing SDS-PAGE (**A**, **C**) and size exclusion chromatography (**B**). **A** SDS-PAGE of the purified sDQ2 preparation. **B** Analysis and fractionation of the produced sDQ2 by size exclusion chromatography. **C** The three fractions obtained in **B** were reanalyzed by SDS-PAGE. Lanes a–c, fractions 1, 2 and 3, respectively (2.2- μ g protein loaded); lane d,

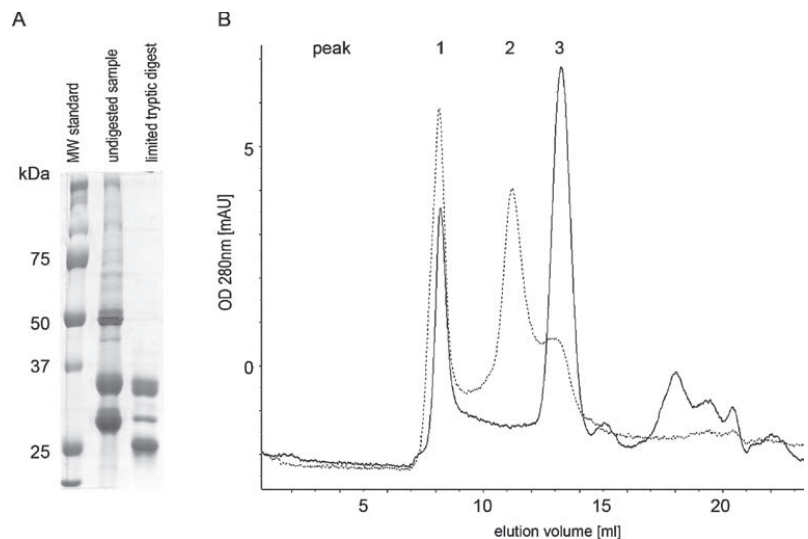
molecular weight standard (BIO-RAD, precision plus protein™ standard); lane e and f, fraction 1 and 2, respectively (29 μ g protein loaded). Amounts of protein loaded onto the gel were calculated from the known amount of total protein subjected to size exclusion chromatography and determination of the peak area

peptide SSIPPEELPHNPLEHDPLHPR as the N terminus of the protein. The observed 50-kDa protein therefore represents a N-terminally truncated derivative of the *D. melanogaster* calcium binding protein (DCB-45), which is described as an elongated homolog of SCF (Kobayashi et al. 1998). The identified protein is highly acidic with a *pI* of 4.27.

Limited digestion of sDQ2/DCB-45 complexes

To investigate the interaction between DCB-45 and sDQ2, we subjected the obtained sDQ2 preparation to limited tryptic digestion with immobilized TPCK trypsin. As evaluated by SDS-PAGE, a digestion time of 17 min at 37°C was found to be optimal to fully degrade DCB-45

Fig. 2 Limited tryptic digestion of the purified sDQ2 preparation analyzed by SDS-PAGE (**A**) and size exclusion chromatography (**B**). **A** Overlay of the size exclusion chromatograms of the nondigested (broken line) and trypsin treated sDQ2 preparation (full line)



(Fig. 2A). Under these conditions, the sDQ2 α -chain remains intact whereas most of the β -chain undergoes a small truncation which does not interfere with formation of a stable and functional heterodimer (see below). Size exclusion chromatography of the same digested sample demonstrated a complete disappearance of the sDQ2/DCB-45 complex which eluted in peak 2 of the untreated sample (Fig. 2B). Notably, the limited proteolysis resulted in a significant increase of the signal intensity observed for peak 3. Following purification of peak 3, the sDQ2 in this preparation was found to be stable for at least 2 days when coincubated at 37°C not only with the high affinity HLA-DQ2 binding peptide P198 but also without adding peptide for at least 24 h at 4°C (data not shown). This suggests that sDQ2 in the partly digested sample remains as a stable heterodimer with an occupied and thereby stabilized peptide binding groove. Peak 3 of the trypsin-treated sDQ2 preparation (Fig. 2B) was concentrated over a 10-kDa spin filter (Vivaspin), washed with PBS, and treated with 0.1% TFA to elute ligands occupying the peptide binding groove. Analysis of the eluted peptides by MALDI-TOF/TOF MS identified four tryptic peptides which were all derived from the 54 N-terminal amino acid sequence of DCB-45 (Fig. 3). These peptides share a 23 amino acid long core sequence (HFDGGEHNAQFDHEAFLGPDESK), suggesting the interaction site is within this region. Besides the DCB-45-derived peptides, we found low signal intensities for several other peptides in the MALDI-TOF spectrum. One of these peptides could be identified as STEFSEDLLEDDLSDIDENEEFLR, and this fragment is derived from the *D. melanogaster* protein RH45818p (Fig. 3).

Peptide binding ability of sDQ2/DCB-45 before and after limited tryptic digestion

The sDQ2/DCB-45 complex (i.e., peak 2 in Fig. 2B, broken line) was purified, and its peptide binding ability was analyzed using the fluorescence-labeled HLA-DQ2 high affinity ligand P198. Peptide binding to sDQ2/DCB-45 was found to be poor, and also limited tryptic digestion of the complex could only slightly increase binding of P198 (Fig. 4). In the presence of HLA-DM, however, binding of P198 was remarkably increased, both for the trypsin treated and untreated sDQ2/DCB-45.

Identification of the HLA-DQ2 peptide binding frame and binding affinity of DCB-45-derived peptides

To further investigate whether the interaction between DCB-45 and sDQ2 is involving the peptide binding groove, we synthesized five 20-mer peptides with overlapping sequences which covered the 54 amino acid long N-terminal DCB-45 sequence (Table 1). The peptide binding affinity was tested in the competition assay using the 125 I-labeled indicator peptide P198 and EBV-DQ2. In this assay, the highest affinity was found for the 20-mer PRHFDGGEHNAQFDHEAFLG. To identify the exact binding frame, three 15-mer peptides covering this 20-mer region were synthesized and tested. The best binding sequence was found to be HNAQFDHEAFLGPDE with an IC₅₀ of 0.13 μ M. The truncated peptide HNAQFDHEAFL also bound with a good affinity (0.77 μ M), suggesting AQFDHEAFL as the 9-mer core region mediating HLA-DQ2 binding (Table 1). The IC₅₀ value of the

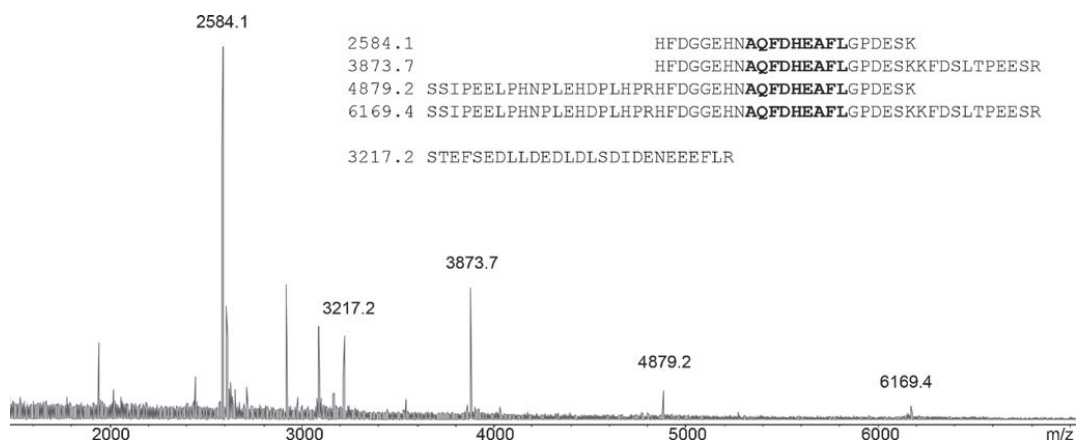


Fig. 3 MALDI-TOF mass spectrum of acid-eluted peptides after trypsin treatment and fractionation of the heterodimeric sDQ2 fraction (peak 3; Fig. 2B). The sequences of the five peptides assigned by their *m/z* values were determined by MALDI-TOF/TOF MS and are given

in the insert. Four of the identified peptides derive from DCB-45. The 9-mer, suggested as the core region interacting with the HLA-DQ2 peptide binding groove, is shown in **bold face**

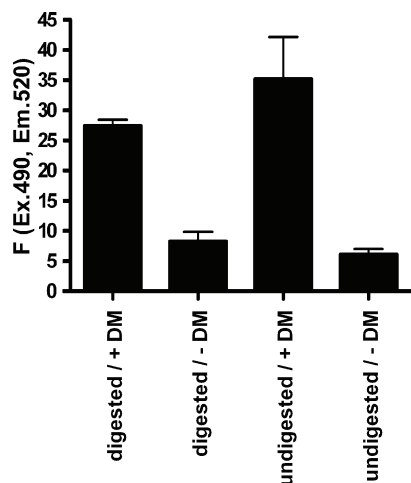


Fig. 4 Binding of the HLA-DQ2 high affinity ligand P198 to sDQ2/DCB-45. Carboxyfluorescein-labeled P198 was incubated with the purified sDQ2/DCB-45 complex with or without prior limited tryptic digestion and in the presence or absence of soluble HLA-DM (sDM). Binding was quantified in a gel-filtration assay and is given by the fluorescence signal (AUC) measured for the sDQ2/DCB-45 complex (undigested sample) or sDQ2 after limited digestion (digested sample). The signals were observed at an elution time of 9.5–9.7 min. Values were obtained from at least two independent experiments; mean and standard errors are shown

54-mer is higher than that of the best binding pentadecapeptide (Table 1), what can obviously not be explained by the slightly lower purity of the 54-mer (72%). Likely, a different secondary structure of the longer peptide hinders the interaction with the peptide binding groove resulting in the observed lower binding affinity.

Peptide binding specificity of sDQ2 compared to EBV-DQ2

The α -I-gliadin epitope and analogs were tested for binding to sDQ2 and EBV-DQ2. For both DQ2 molecules, the calculated relative binding capacity and the ranking of the measured IC₅₀ values for the set of peptides was found to be nearly identical. This strongly indicates that the peptide binding specificity of sDQ2 matches that of EBV-DQ2 (Fig. 5).

Discussion

In this study, we show that soluble HLA-DQ2 molecules without a covalently tethered peptide ligand copurify with the *D. melanogaster*-derived protein DCB-45, when expressed in S2 insect cells. We identified the fragment of DCB-45 which mediates the strong interaction with sDQ2 and obtained evidence indicating that a part of this fragment is binding to the sDQ2 binding groove with high affinity. The sDQ2 stabilizing effect of the DCB-45 protein motivated us to unravel the basis for this interaction as

Table 1 Binding of synthetic peptides to EBV-DQ2 given as IC₅₀ values

| Peptides | IC ₅₀ [μ M] |
|--|-----------------------------|
| SSIPEELPHNPLEHDP LHPRHFDGGGEHNA QFDHEAFL GPDESKKFDLSLTPEESR | 0.8 ^a |
| SSIPEELPHNPLEHDP LHPR | 51.1 ^b |
| NPLEHDP LHPRHFDGGGEHNA | >167 ^b |
| PRHFDGGGEHNA QFDHEAFL G | 0.56 ^b |
| NA QFDHEAFL GPDESKKFD | 1.09 ^b |
| LGPDESKKFDLSLTPEESRRR | 7.76 ^b |
| LHPRHFDGGGEHNA QF | >167 ^b |
| FDGGGEHNA QFDHEAF | 4.64 ^b |
| HNA QFDHEAFL GPDE | 0.13 ^b |
| HNA QFDHEAFL | 0.77 ^b |
| KPLLIIAEDVGEY | 0.09 ^b |
| QLQPFPPQPELPY | 1.1 ^a |
| PQPELPYPQPQLPY | 7.4 ^a |
| KPLLIIAEDVGEY | 0.06 ^a |

IC₅₀ is the peptide concentration required to give 50% inhibition of the binding of the indicator peptide (P198). The IC₅₀ values for the peptides presented in this table were analyzed in two separate sets of experiments. The results from one of three consistent titrations are given. As a reference peptide, P198 (KPLLIIAEDVGEY) was included. The HLA-DQ2 binding frames are indicated in bold face

^a In the first set of experiments, the binding affinity of the 54-mer peptide (72% pure) compared to gliadin-derived peptides P1269 (HLA-DQ2- α -I epitope; QLQPFPPQPELPY), P1274 (HLA-DQ2- α -II epitope PQPELPYPQPQLPY) and indicator peptide P198 was analyzed

^b In the second set of experiments, the binding of overlapping synthetic peptides spanning the 54 amino acid long N-terminal region of DCB-45 was analyzed

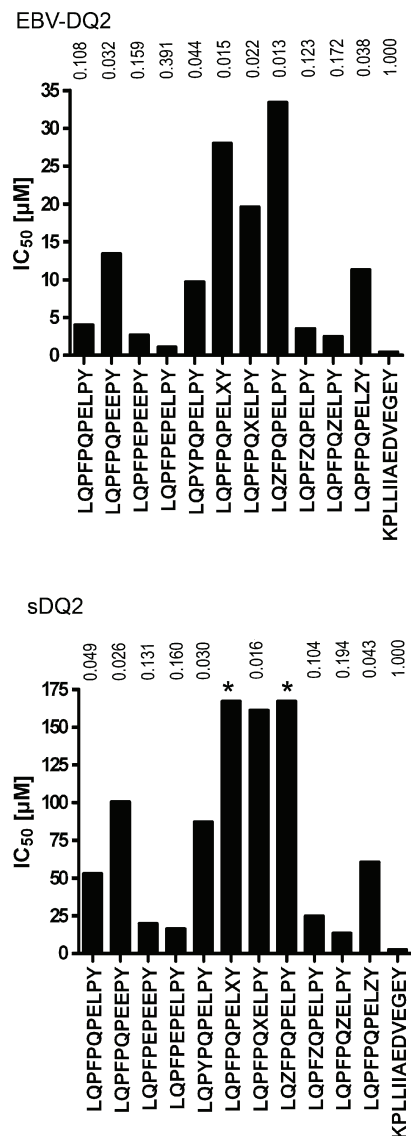


Fig. 5 Peptide binding specificity of sDQ2 compared to EBV-DQ2. sDQ2 (1 μM) and EBV-DQ2 (0.3 μM) was incubated with the radioactively labeled indicator peptide P198 (KPLLIIAEDVEGEY) and titrations of competitor peptides. The HLA-DQ2-α-I gliadin epitope (LQPFQPELPY) and analogs were used. The bars show the IC₅₀ values (in μM). The relative binding capacity for each peptide is given and was calculated as the ratio of the IC₅₀ of the reference peptide P198 to the IC₅₀ of the test peptide. The Z in the amino acid sequence denotes norvaline, X denotes 4-hydroxyproline. All peptides marked with an asterisk (*) have an IC₅₀ value higher than 167 μM and the exact value could not be determined

part of an effort to design compounds that can block HLA-DQ2-mediated antigen presentation.

Expression of HLA class II molecules in insect cells is a commonly used method for the production of recombinant molecules (De Wall et al. 2006; Stern and Wiley 1992; Wallny et al. 1995). Initially, DR1 and other MHC class II molecules were expressed without a covalently tethered peptide as so called “empty” class II molecules (Stern and Wiley 1992). The expression of recombinant molecules with a high affinity ligand covalently tethered to the N terminus of the β-chain (Crawford et al. 1998), however, gave improved expression and stable molecules and this strategy thus developed as a preferred expression modality. Our early work to express sDQ2 without a covalently tethered peptide in the S2 cell system gave unusually high protein yields and stable molecules what is in contrast to a poor expression yield and unstable molecules obtained with similar baculoviral constructs in Sf9 cells (Quarsten et al., unpublished). When analyzing and comparing peptide binding specificity of these sDQ2 molecules with detergent-solubilized HLA-DQ2 purified from EBV-transformed B cells (EBV-DQ2), we found the same peptide binding preferences but approximately a tenfold increase of the IC₅₀ values for each peptide when using the sDQ2 molecules (Fig. 5).

In mammalian antigen presenting cells the assembly of HLA class II α- and β-chains is dependent on the invariant chain (Ii) protein. In the ER, Ii is protecting and stabilizing the class II binding groove with its class II-associated invariant chain peptide (CLIP) sequence, and the Ii is directing the class II molecules to the late endosomes where HLA-DM catalyzes the exchange of CLIP for antigenic peptides (Busch et al. 2005; Cresswell 1994). In S2 cells, this mechanism for sDQ2 assembly and loading is missing. One might thus expect that the unoccupied amphiphilic peptide binding groove is rendering the class II molecules unstable and prone to aggregation (Rabinowitz et al. 1998; Vogt et al. 1997). For this reason, it has been argued that so-called “empty” class II molecules are not truly empty but filled with loosely bound peptides. We thus wanted to characterize sDQ2 molecules produced in S2 cells in more detail.

By SDS-PAGE analysis under reducing conditions, we observed a strong band at 50 kDa in addition to the bands of the sDQ2 α- and β-chains. A similar type of contamination has not been observed with other soluble HLA molecules produced in baculovirus infected Sf9 insect cells, nor in the wild type EBV-DQ2 produced by EBV-transformed B cells. This contaminating protein was identified as the *D. melanogaster*-derived DCB-45 protein and our findings suggest a tight interaction between sDQ2 and DCB-45. The complex does not dissociate under the condition of high pH and high salt concentrations used for elution from the

antibody affinity column (pH 11.5 and 2 M Tris), and treatment at low pH resulted in precipitation rather in dissociation of the two proteins. In contrast, the sDQ2 molecules found in the large aggregates (>600 kDa) apparently are not associated with DCB-45, suggesting that sDQ2 is unstable in the absence of DCB-45. The sDQ2/DCB-45 interaction thus likely explains the observed high expression yield of peptide-receptive sDQ2.

DCB-45 is a highly acidic protein (*pI* 4.27) and HLA-DQ2 has a strong preference for binding negatively charged residues in several of its anchor positions (Kim et al. 2004). We therefore investigated whether the sDQ2/DCB-45 binding is also involving the peptide binding groove of sDQ2. The characterization of such a strong interaction could be helpful to design new lead structures for high affinity HLA-DQ2 blockers. Such reagents have been discussed in the treatment of celiac disease to inhibit HLA-DQ2-restricted presentation of gluten derived epitopes (Bergseng et al. 2005; Xia et al. 2006, 2007).

In the presence of HLA-DM, binding of the known HLA-DQ2 ligand P198 to the isolated sDQ2/DCB-45 complex increased. The same effect of HLA-DM was observed after limited proteolysis of this complex leading to almost exclusive degradation of DCB-45 and a markedly increase of heterodimeric sDQ2 (Fig. 2). Analysis by size exclusion chromatography showed that this fraction was stable. Interestingly, DCB-45-derived peptides could be released at low pH and were found to originate from the N-terminal 54-mer (Fig. 3). As these findings strongly indicate a specific interaction of this region with the sDQ2 peptide binding groove, a more detailed mapping was performed. Using detergent-solubilized EBV-DQ2, the pentadecapeptide HNAQFDHEAFLGPDE was found as the best binder with an affinity similar to that of the P198 ligand. The data also suggest the sequence AQFDHEAFL as the 9-mer core region with negative charges in positions P4 and P6 and a large hydrophobic residue in position P9 that are favorable for interaction with the sDQ2 peptide binding groove. This sequence contains alanine in position 1, which is not expected to be favorable due to its small size and the preference that HLA-DQ2 has for bulky hydrophobic anchor residues at this position (van de Wal et al. 1996; Vartdal et al. 1996). Notably, peptide binding to HLA-DQ2 is characterized by contributions from anchor residues in position 1, 4, 6, 7, and 9 (Kim et al. 2004). This is in contrast to peptide binding to HLA-DR molecules which has a dominant contribution of the P1 anchor (Jardetzky et al. 1990; Stern et al. 1994). Given that the peptide AQFDHEAFL has optimal anchor residues at positions P4 (D), P6 (E), and P9 (L), an alanine residue at P1 with no repulsive effect may well be compatible with the high affinity binding of this ligand. Moreover, *in silico* searching for possible binding registers based on binding

studies with peptide libraries (Jüse et al., unpublished) gave no better candidates for high affinity HLA-DQ2 ligands within the 54-mer region. The DCB-45-derived 15-mer HNAQFDHEAFLGPDE is a good HLA-DQ2 binder, and we found that it is binding to HLA-DQ2 with a higher affinity than the HLA-DQ2- α -I-gliadin peptide that is an immunodominant epitope of celiac lesion derived CD4 T cells. Work parallel to this study (Xia et al. 2007) has demonstrated that even higher binding affinity than that of the DCB-45-derived pentadecapeptide as well as high proteolytic stability are required to obtain HLA-DQ2 blockers that effectively prevent activation of gliadin reactive T cells. Several chemical modifications that improve HLA-DQ2 binding affinity and proteolytic resistance would thus be required to make the DCB-45-derived ligand becoming an effective HLA-DQ2 blocker.

The observations presented in this paper are relevant for anyone aiming to express MHC class II molecules in S2 cells. DCB-45 is influencing the availability of free and functional peptide binding site of recombinant sDQ2, resulting in a decreased peptide binding capacity of the expressed molecules. Our findings illustrate that *de novo* interaction of recombinant MHC class II molecules with proteins derived from the expression system can occur. Prior to large scale use in peptide binding studies, such MHC class II molecules should be carefully characterized and compared with respect to peptide binding specificity and capacity to the wild type molecules, which are, e.g., purified from Epstein–Barr virus-transformed B lymphoblastoid cell lines.

Acknowledgements This work has been funded as part of a European Commission Marie Curie Research Training Network (MRTN-CT-2004-512385). We would like to thank Elizabeth D. Mellins, Stanford University, for providing S2 cells producing HLA-DM and Chu-Young Kim, Stanford University, for synthesizing some of the peptides.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bergseng E, Xia J, Kim CY, Khosla C, Sollid LM (2005) Main chain hydrogen bond interactions in the binding of proline-rich gluten peptides to the celiac disease-associated HLA-DQ2 molecule. *J Biol Chem* 280:21791–21796 doi:10.1074/jbc.M501558200
- Busch R, Rinderknecht CH, Roh S, Lee AW, Harding JJ, Burster T, Homell TM, Mellins ED (2005) Achieving stability through editing and chaperoning: regulation of MHC class II peptide binding and expression. *Immunol Rev* 207:242–260 doi:10.1111/j.0105-2896.2005.00306.x

- Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO (1995) Receptor-ligand interactions measured by an improved spun column chromatography technique. A high efficiency and high throughput size separation method. *Biochim Biophys Acta* 1243:453–460
- Crawford F, Kozono H, White J, Marrack P, Kappler J (1998) Detection of antigen-specific T cells with multivalent soluble class II MHC covalent peptide complexes. *Immunity* 8:675–682 doi:10.1016/S1074-7613(00)80572-5
- Cresswell P (1994) Assembly, transport, and function of MHC class II molecules. *Annu Rev Immunol* 12:259–293 doi:10.1146/annurev.iy.12.040194.001355
- De Wall SL, Painter C, Stone JD, Bandaranayake R, Wiley DC, Mitchison TJ, Stern LJ, Decker BS (2006) Noble metals strip peptides from class II MHC proteins 1. *Nat Chem Biol* 2:197–201 doi:10.1038/nchembio773
- Fleckenstein B, Qiao SW, Larsen MR, Jung G, Roepstorff P, Sollid LM (2004) Molecular characterization of covalent complexes between tissue transglutaminase and gliadin peptides. *J Biol Chem* 279:17607–17616 doi:10.1074/jbc.M310198200
- Greenwood FC, Hunter WM, Glover JS (1963) The preparation of I-131-labelled human growth hormone of high specific radioactivity. *Biochem J* 89:114–123
- Jardetzky TS, Gorga JC, Busch R, Rothbard J, Strominger JL, Wiley DC (1990) Peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. *EMBO J* 9:1797–1803
- Johansen BH, Buus S, Vartdal F, Viken H, Eriksen JA, Thorsby E, Sollid LM (1994) Binding of peptides to HLA-DQ molecules: peptide binding properties of the disease-associated HLA-DQ (alpha 1*0501, beta 1*0201) molecule. *Int Immunol* 6:453–461 doi:10.1093/intimm/6.3.453
- Kaiser E, Colescott RL, Bossinger CD, Cook PI (1970) Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem* 34:595–598 doi:10.1016/0003-2697(70)90146-6
- Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM (2004) Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci USA* 101:4175–4179 doi:10.1073/pnas.0306885101
- Kobayashi M, Aita N, Hayashi S, Okada K, Ohta T, Hirose S (1998) DNA supercoiling factor localizes to puffs on polytene chromosomes in *Drosophila melanogaster*. *Mol Cell Biol* 18:6737–6744
- Quarsten H, McAdam SN, Jensen T, rentz-Hansen H, Molberg O, Lundin KE, Sollid LM (2001) Staining of celiac disease-relevant T cells by peptide-DQ2 multimers. *J Immunol* 167:4861–4868
- Rabinowitz JD, Vrljic M, Kasson PM, Liang MN, Busch R, Boniface JJ, Davis MM, McConnell HM (1998) Formation of a highly peptide-receptive state of class II MHC. *Immunity* 9:699–709 doi:10.1016/S1074-7613(00)80667-6
- Sloan VS, Cameron P, Porter G, Gammon M, Amaya M, Mellins E, Zaller DM (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* 375:802–806 doi:10.1038/375802a0
- Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* 169:345–350 doi:10.1084/jem.169.1.345
- Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215–221 doi:10.1038/368215a0
- Stern LJ, Wiley DC (1992) The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. *Cell* 68:465–477 doi:10.1016/0092-8674(92)90184-E
- Thorsby E, Ronningen KS (1993) Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 (insulin-dependent) diabetes mellitus 4. *Diabetologia* 36:371–377 doi:10.1007/BF00402270
- Todd JA, Bell JI, McDevitt HO (1987) HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 329:599–604 doi:10.1038/329599a0
- Van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Koning F (1996) Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. *Immunogenetics* 44:246–253 doi:10.1007/BF02602553
- Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanovic S, Eriksen JE, Sletten K, Thorsby E, Rammensee HG, Sollid LM (1996) The peptide binding motif of the disease associated HLA-DQ (alpha 1*0501, beta 1*0201) molecule. *Eur J Immunol* 26:2764–2772 doi:10.1002/eji.1830261132
- Vogt AB, Moldenhauer G, Hammerling GJ, Kropshofer H (1997) HLA-DM stabilizes empty HLA-DR molecules in a chaperone-like fashion. *Immunol Lett* 57:209–211 doi:10.1016/S0165-2478(97)00061-8
- Wallny HJ, Sollami G, Karjalainen K (1995) Soluble mouse major histocompatibility complex class II molecules produced in *Drosophila* cells. *Eur J Immunol* 25:1262–1266 doi:10.1002/eji.1830250520
- Xia J, Bergseng E, Fleckenstein B, Siegel M, Kim CY, Khosla C, Sollid LM (2007) Cyclic and dimeric gluten peptide analogues inhibiting DQ2-mediated antigen presentation in celiac disease. *Bioorg Med Chem* 15:6565–6573 doi:10.1016/j.bmc.2007.07.001
- Xia J, Siegel M, Bergseng E, Sollid LM, Khosla C (2006) Inhibition of HLA-DQ2-mediated antigen presentation by analogues of a high affinity 33-residue peptide from alpha2-gliadin. *J Am Chem Soc* 128:1859–1867 doi:10.1021/ja056423o

**Design of new high-affinity peptide ligands for human leukocyte antigen-DQ2
using a positional scanning peptide library**

Ulrike Jüse, Yvonne van de Wal, Frits Koning, Ludvig M. Sollid, Burkhard
Fleckenstein

Hum Immunol. 2010 May;71(5):475-81

Paper III

Assessing high affinity binding to HLA-DQ2.5 by a novel library-based approach

Ulrike Jüse, Magnus Arntzen, Peter Højrup, Burkhard Fleckenstein and Ludvig M. Sollid

Manuscript (submitted to Bioorganic & Medical Chemistry, 03/2010)

Assessing high affinity binding to HLA-DQ2.5 by a novel peptide library-based approach

Ulrike Jüse^{a*}, Magnus Arntzen^{b,c,d}, Peter Højrup^e, Burkhard Fleckenstein^f and Ludvig M. Sollid^{a,f}

^aCentre for Immune Regulation, Institute of Immunology, Oslo University Hospital-Rikshospitalet, 0027 Oslo, Norway

^bProteomics Core Facility, Oslo University Hospital-Rikshospitalet & University of Oslo, 0027 Oslo, Norway

^cThe Biotechnology Centre of Oslo, University of Oslo, 0317 Oslo, Norway

^dProteomics Core Facility, Norwegian University of Life Sciences, 1432 Ås, Norway

^eInstitute of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense, Denmark

^fCentre for Immune Regulation, Institute of Immunology, University of Oslo, 0027 Oslo, Norway

*Corresponding author: Tel.: ++47 23 07 30 17; fax: ++47 23 07 35 10; e-mail:

ulrike.juse@rr-research.no

Abstract:

Here we report on a novel peptide library based method for HLA class II binding motif identification. The approach is based on water soluble HLA class II molecules and soluble dedicated peptide libraries. A high number of different synthetic peptides are competing to interact with a limited amount of HLA molecules, giving a selective force in the binding. The peptide libraries can be designed so that the sequence length, the alignment of binding registers, the numbers and composition of random positions are controlled, and also non-natural amino acids can be included. Selected library peptides bound to HLA are then isolated by size exclusion chromatography and sequenced by tandem mass spectrometry online coupled to liquid chromatography. The MS/MS data are subsequently searched against a library defined database using a search engine such as Mascot, followed by manual inspection of the results. We used two dodecamer and two decamer peptide libraries and HLA-DQ2.5 to test possibilities and limits of this method. The selected sequences which we identified in the fraction eluted from HLA-DQ2.5 showed a higher average of their predicted binding affinity values compared to the original peptide library. The eluted sequences fit very well with the previously described HLA-DQ2.5 peptide binding motif. This novel method allows the analysis of several thousand synthetic sequences concomitantly in a simple water soluble format.

Keywords:

Dedicated peptide libraries

HLA-DQ2.5

Mass spectrometry

Peptide blockers

Celiac disease

1. Introduction

HLA associations are seen for most autoimmune diseases. HLA-DQ2.5 (DQA1*0501/DQB1*0201) of the B8-DR3-DQ2 haplotype is associated with particularly many diseases, including type 1 diabetes, Addison's disease, systemic lupus erythematosus and celiac disease.^{1,2,3} Celiac disease is a complex inflammatory disorder of the small intestine, and HLA-DQ2.5 exerts a central role in the pathogenesis of this disease by presenting gluten peptides to specific CD4+T-cells in the intestinal mucosa.⁴ Over 90% of the celiac disease patients express HLA-DQ2.5.^{5,6} For celiac disease, but also other HLA-DQ2.5 associated diseases, specific interference with this key molecule can represent a novel treatment strategy. An HLA blocker should interfere with the binding of immunogenic peptide(s), but should not be recognized by any T cell receptor. So far, the most promising approaches for blocking of the HLA-DQ2.5 mediated presentation of gluten derived antigens is the use of cyclic and dimeric gluten peptides,⁷ or the introduction of a large, chemically modified side chain into gluten epitopes.⁸ However, the efficacy of the designed blockers for inhibiting T cell activation *in vitro* is inadequate, and it appears that insufficient affinity of the blockers for HLA-DQ2.5 is at least one limitation.

In this study we present a method to identify optimal HLA class II ligands. By incubating peptide libraries with several randomized positions with limited amounts of HLA molecules, an environment of strong competition is created in which HLA-DQ2.5 selects for the best binders (Fig. 1). Identification of the selected ligands is done by mass spectrometry, and thus also non-natural amino acids can be included in the libraries. We here report on the strengths and weaknesses of the method and present data on HLA-DQ2.5 ligands and the peptide binding motif of this molecule.

2. Results

2.1. Peptide libraries with a single randomized position – libraries A and B

Peptide libraries with a single randomized position were used to optimize separation of HLA-DQ2.5-peptide complexes from unbound peptides and to explore the possibilities of this novel method. Due to the low complexity, the libraries could be easily analyzed by MALDI-TOF mass spectrometry without the need for LC-based peptide separation. In addition, as each peptide sequence is defined by its $[M+H]^+$ value, sequence identification of the detected peptides by database searching was not required. Notably, for some peptides the isotopic envelopes overlap and the contribution of the different peptides to signal intensity was calculated by using their theoretical isotopic envelope (supplementary Fig. S1). We used two dodecapeptide libraries each carrying a single X position in the anchor positions P1 or P9 (Table 1). The mass spectra of library A, which were acquired from the original peptide library and from the peptides eluted from HLA-DQ2.5, are shown in Fig. 2. All $[M+H]^+$ values calculated from the format of this library were observed in the mass spectrum of the original library. For the eluted sample, significant signals ($18 > x$ for Y) were only obtained for $[M+H]^+$ values which corresponded to sequences with bulky hydrophobic amino acids in position P9 (V, L/I, F, Y, W). Library B, carrying the randomized position P1, was analyzed by the same method. For the original library, $[M+H]^+$ values corresponding to all expected peptide sequences, except the one with proline in the X position, were identified. By contrast to library A, the effect of the selection by HLA-DQ2.5 was less pronounced for library B as we found signals for all 19 peptide sequences in the eluted sample (data not shown).

2.2. Peptide library with four dedicated Z positions – library C

LC-MS/MS analysis of peptide library C on an LTQ-Orbitrap XL mass spectrometer resulted in 5059 MS/MS events for the original library (Table 2). A Mascot search of these MS/MS events against a library defined database led to the identification of 669 peptides (Table 2). The RD values for each of the amino acids in the dedicated Z positions of these peptides were in the range of 0.5 and 1.5, not far from the expected RD value of 1 (Fig. 3A).

LC-MS/MS analysis of the sample after elution from HLA-DQ2.5 resulted in 707 MS/MS events (Table 2). Mascot reported 81 unique top ranking sequences, however, for the majority of these 81 MS/MS events, more than one peptide sequence was assigned by a significant score. Therefore, all of these MS/MS events were subjected to additional manual data inspection, including examination of their retention times and elution profiles, as isobaric peptides can differ slightly in their LC retention times (Fig. 4). Additionally, for these 81 MS/MS events, we manually compared the observed fragment ions with their theoretical fragmentation patterns. This manual data evaluation resulted in the exclusion of 12 sequences and inclusion of 12 sequences. Thus, 81 unique sequences were identified from library C after selection by HLA-DQ2.5 (supplementary Table S1).

Based on these final 81 identified sequences, RD values were calculated for each amino acid present in the four dedicated Z positions. These RD values indicated that HLA-DQ2.5 favored valine, alanine and glutamate at position P6, glutamate in position P7 and P8 and valine, leucine, glutamate, tyrosine, aspartate, methionine, phenylalanine and tryptophan at position P9 (Fig. 3B). In contrast, a few amino acids

were not detected or present in low amounts in the dedicated Z positions. These were glutamine, threonine, asparagine and glycine at position P6, glycine at position P7 and glycine, serine and proline at position P9 (Fig. 3B).

For all peptides theoretically present in the original library C, and for the 81 peptides identified in the eluted sample, we calculated the sum of the *relC* values ($\sum relC$). For the eluted peptides the average of these $\sum relC$ values was increased by 1.7 compared to the average of the $\sum relC$ values of all peptides of library C (Fig. 5). This finding suggests that HLA-DQ2.5 has indeed selected for optimal HLA-DQ2.5 ligands.

2.3. Peptide library with six dedicated Z positions – library D

LC-MS/MS analysis of peptide library D resulted in 5791 MS/MS events for the original library and 1212 MS/MS events for the eluted peptide sample after incubation with HLA-DQ2.5 (Table 2). The MS/MS data obtained for the original library were subjected to a Mascot search, and we could identify 1985 of the 65350 theoretically possible peptide sequences (Table 2). Based on these peptides, we calculated RD values for the amino acids in the dedicated Z positions, and found that they varied between 0.1 and 3.8 (Fig. 6A).

After peptide selection by HLA-DQ2.5, the obtained MS/MS data were searched by Mascot resulting in the identification of 179 unique top ranking peptide sequences. All MS/MS events that had more than one peptide sequence assigned by a significant score were manually inspected. As only 3% of the theoretical sequences of library D were identified in the sample before incubation with HLA-DQ2.5, this manual evaluation did not include comparison of retention times and elution profiles but

consisted only of comparison of the observed with the theoretical fragmentation patterns. This manual data evaluation increased the number of identified sequences to 261 (Table 2, supplementary Table S2). The average of the $\sum relC$ values of the eluted peptides was increased by 3.3 compared to the average of the $\sum relC$ values of all peptides present in the original library D (Fig. 5). This suggests an enrichment of optimal HLA-DQ2.5 binding peptides from library D also.

As RD values of amino acids in the dedicated positions of the original library D showed substantial deviation from 1, there is a likely bias in the RD values calculated for the HLA-DQ2.5 selected sample (Fig. 6B). In order to compensate for this bias, we calculated adjusted RD values by dividing the RD values obtained for the HLA-DQ2.5 selected sample by the RD values obtained for the original library (Fig. 6C). These adjusted RD values suggest that HLA-DQ2.5 favored phenylalanine and tryptophane at position P1; valine, glutamate and leucine at position P4; valine, glutamate, proline, alanine and leucine at position P6; glutamate and leucine at position P7, glutamate at position P8, and phenylalanine, leucine, tryptophan, tyrosine and methionine at position P9 (Fig. 6C). Several amino acids were not detectable or present at low abundance at some positions, when the peptide sample selected by HLA-DQ2.5 was analyzed. These included proline, valine and glutamate at position P1; threonine, glycine, tryptophane and tyrosine at position P4, glutamine at position P6; glycine at P7; and glycine, glutamine, serine, valine and glutamate at position P9.

3. Discussion and conclusions

Here we report on a new method for the specific enrichment of peptides with high affinity to a HLA class II molecule from complex soluble dedicated peptide libraries. This method allows identification of high affinity ligands as well as refined definition of peptide binding motifs.

Interfering with antigen presentation by blocking the binding groove of HLA class II molecules is a conceivable strategy for the treatment of HLA associated diseases.

HLA-DQ2.5 is a particular attractive target, as many diseases show associations with this HLA molecule.⁹ Development of HLA-DQ2.5 specific blockers has however so far given limited results, and an obvious limitation is the HLA binding affinity of the blockers. Thus there is a need for tools that allow the identification of high affinity ligands and that allow for testing of compounds incorporating non-natural amino acids. Here we present a peptide library based method, that should work equally well with natural as well as non-natural amino acids, as the detection is undertaken by mass spectrometry. The proof of principle of the method has been established by use of natural amino acids and dedicated peptide libraries.

We first analyzed peptide libraries with a single fully randomized X position in the anchor positions P1 and P9. By comparing the MALDI-TOF mass spectra of the peptide library before incubation with the spectra after elution, we found a specific depletion of several peptides. Peptides with bulky hydrophobic amino acids at position P9 were favored by HLA-DQ2.5 and were the only sequences observed in the eluted sample. At position P1 however, HLA-DQ2.5 was less selective and bound all of the 19 tested peptides. Prior studies have shown, that HLA-DQ2.5 both at

positions P9 and P1 prefer amino acids with bulky hydrophobic side chains.^{10,11}

However, many immunogenic gliadin peptides have the non-bulky amino acid proline at position P1, but never in position P9.^{12,13,14} This suggests that at position P1, a non-preferred amino acid can be selected, provided all other anchors are optimal.

Next we designed two dedicated peptide libraries with four and six dedicated positions (Z). For peptide library C, with four dedicated Z positions, we chose a decapeptide library format with four neighboring dedicated Z positions in the C-terminal part. Notably, the N-terminal anchor positions P1 and P4 were occupied by alanine, a non-optimal anchor residue for HLA-DQ2.5. By using this format, we sought to solely identify sequences with optimal residues in all of the three anchor positions P6, P7 and P9. Each of the dedicated Z positions was composed of 2 to 13 amino acids, that previously have been demonstrated to be intermediate or good residues in the corresponding positions (Fig. 7).¹⁵ In peptide library D, two additional dedicated Z positions were placed in the two remaining anchor positions, P1 and P4. These Z positions were composed of altogether five (at P1) and nine (at P9) amino acids, that previously have been demonstrated to be intermediate and good anchor residues for these positions (Fig. 7). The two additional Z positions increased the complexity of the peptide library by 50-fold compared to library C. In library D, a total of 65 000 sequences competed for binding to HLA-DQ2.5. Due to the complexity it was necessary to separate peptides in libraries C and D by liquid chromatography prior to MS/MS analysis, and for sequence identification a library-defined database and a search engine (Mascot) was required.

The HLA-DQ2.5 selected sequences identified from both the C and D libraries had higher average $\sum relC$ values than the peptides each library was composed of. This is suggesting that the HLA-DQ2.5 affinity matrix did select for ligands with increased affinity in both cases. From library C we identified a panel of 81 peptide ligands (Table 2, supplementary Table S1), and we established RD values which designated preferred and less preferred anchor residues at the P6, P7 and P9 positions (Fig. 3B). Notably, about half of the theoretically possible sequences were identified when analyzing the original library C. The approximately equal distribution of amino acids in the dedicated positions, argues for a high quality of the peptide libraries used in this study.

The data obtained from library D is less informative. There was an uneven representation of amino acids in the dedicated positions of the original library which was probably introduced during the synthesis of the library, or which is due to a random error, introduced when Mascot is searching a database of that high complexity. The list of identified peptides from the sample selected by HLA-DQ2.5 might therefore also be biased. We therefore adjusted the RD values for these selected peptides, and these adjusted RD values give some insight into the HLA-DQ2.5 binding motif.

Our approach can easily be extended to the incorporation of non-natural amino acids into the peptide libraries, as the identification of selected ligands is done by tandem mass spectrometry. It appears that efficient HLA-DQ2.5 blockers can not solely be made of natural amino acids.⁷ Ongoing research has started exploring high affinity binders to HLA-DQ2.5 with both proteinogenic and modified, non-natural amino

acids.^{7,8} Hopefully this work can lead to the development of blockers which show sufficient efficacy in vitro and in vivo.

4. Experimental

4.1. Soluble HLA-DQ2.5 molecules, expression and purification

Water soluble HLA-DQ2.5 molecules (DQA1*0501/DQB1*0201) tethered with the DQ2- α -I-gliadin epitope (sDQ2- α I) were expressed and purified as described before.¹⁶ The purity of HLA-DQ2.5 preparations was analyzed by SDS-PAGE. Protein concentrations were determined by a BCA protein assay kit (Pierce, Rockford, IL, USA).

4.2. Peptide library design

We tested altogether four peptide libraries based on the same sequence scaffold (Table 1). Two dodecapeptide libraries, named A and B, were randomized at one position each (position X) which was composed of an equimolar mixture of all natural amino acids, except cysteine. These X positions were either placed at anchor position P9 (library A) or anchor position P1 (library B) of the HLA-DQ2.5 binding register (Table 1). Further, we used two acetylated decapeptide amide libraries with 4 and 6 dedicated positions with limited variation. In these libraries the dedicated positions (positions Z) were composed of an equimolar mixture of 2 to 13 different proteinogenic amino acids and placed in several anchor positions and one non-anchor position of the HLA-DQ2.5 binding frame (Fig. 1, Table 1). In library C, the Z positions were placed in positions P6, P7, P8 and P9, and in library D, the Z positions were localized in positions P1, P4, P6, P7, P8 and P9. The amino acids selected for the Z positions of the libraries C and D were residues assumed to be intermediate to optimal anchors based on their relative competition (*relC*) values (>1.5) as previously established by a scanning peptide library approach (Fig. 7).¹⁵ Further, both libraries

carried lysine in position P5 and arginine in position P10 to prevent binding of each peptide in multiple binding registers. In all other positions alanine was used.

4.3. Peptide library synthesis

Peptide libraries with one X position were synthesized on Rink amide MBHA-resin using Fmoc/HBTU chemistry and a pipetting robot (Syrro I, MultiSynTech, Bochum, Germany). The randomized X position was coupled by using an equimolar mixture of all proteinogenic amino acids except cysteine. Coupling was performed twice using equimolar amounts of amino acids and available amino groups. Libraries C and D were purchased from EZBiolab (EZBiolab Inc., Carmel, USA).

4.4. Competitive binding conditions

sDQ2- α I molecules were treated with thrombin (0.1 U per 4.6 μ g of sDQ2- α I) for 2 hours at room temperature, followed by a treatment with 0.5 μ g 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride. Subsequent, 50 μ g peptide library and 1.5 μ g soluble HLA-DM (a gift from E. Mellins) were added to 8.3 μ g thrombin-cleaved sDQ2- α I in a total volume of 80 μ l phosphate buffered saline containing 22.5 mM citric acid and 50.5 mM Na₂HPO₄ (pH 5.3). The sample was incubated for 48 hours at 37 °C.

4.5. Separation of high affinity peptides by size exclusion chromatography

Library derived peptides bound to sDQ2 were separated from unbound peptides by preparative size exclusion chromatography on an Äkta purifier system (Amersham Bioscience Corp., Piscataway, NJ, USA) using a Superdex 200 10/300 GL column (Amersham Bioscience) and 15 mM phosphate buffered saline as the mobile phase.

The sDQ2 containing fraction was concentrated and sDQ2-bound peptides were eluted by adjusting to pH 1 with trifluoroacetic acid. The samples were incubated for 30 min and peptides were purified on a reversed phase (Poros 20 R2, Applied Biosystems, Foster City, CA, USA) packed in 20- μ l GELoader tips (Eppendorf, Hamburg, Germany). Peptides were eluted with 10 μ l 99.9% acetonitrile and the sample was dried.

4.6. Identification of peptides from libraries with a single X position by MALDI-TOF mass spectrometry

A volume of 1 μ l of the eluted peptide samples was applied on a stainless steel MALDI target plate and 0.5 μ l of 10 g/L α -cyano-4-hydroxycinnamic acid in 70% acetonitrile/0.1% TFA was added. MALDI-TOF mass spectra were acquired on an Ultraflex II instrument (Bruker Daltonics, Bremen, Germany) and peptide sequences were identified manually. For all peptides with an overlapping isotopic envelope, the theoretical isotopic distribution was taken into consideration.

4.7. LC-MS/MS analysis of samples from dedicated libraries

The dried peptide samples were dissolved in 0.1% formic acid and analyzed by nanoLC-MS/MS using an EASY nLC system (Proxeon A/S, Odense, Denmark) coupled to a LTQ-Orbitrap XL mass spectrometer (ThermoScientific, Bremen, Germany) equipped with a nano-electrospray ion source. The peptides were separated on a 100 micron ID capillary column 18 cm length, packed with Reprosil C₁₈, 3 micron. The instrument was set to a linear gradient of 0-34% B for 50 minutes (solvent B was 95% acetonitrile), the threshold for fragmentation was set to 15.000

and the range was 300-1800. Of the original library samples, 0.4 µg were analyzed by using the same settings.

4.8. Peptide identification after analysis of samples from dedicated libraries

Obtained LC-MS/MS data were extracted from the Orbitrap raw files using the Proteome Discoverer software (Thermo Fisher Scientific Inc., Waltham, MA, USA) with default settings and searched against a library defined database using Proteome Discoverer and the Mascot search engine (Matrixscience Ltd, Boston, MA, USA). The following settings were selected: amidated C-terminus and acetylated N-terminus as fixed modifications, Arg-C as the enzyme, no missed cleavages, peptide mass tolerance of 10 ppm, and fragment ion mass tolerance of 0.6 Da. Search results for the original library and for the peptide sample selected by HLA-DQ2.5 were copied to an Excel spreadsheet giving information on the peptide sequences reported for each MS/MS event, together with their ion scores, rank order and retention times. Due to the format of the library, Mascot assigned several isobaric peptides with high scores for most of the MS/MS events. As these isobaric peptides share many of their fragment ion masses, the exact sequence of a selected parent ion can often only be differentiated on the basis of one or a few y- and/or b-fragments. If more than one hit was reported by Mascot for a given MS/MS event in the eluted samples, manual data inspection was performed. For both dedicated libraries, the theoretical fragmentation pattern of all hits reported by Mascot for a particular MS/MS event were compared to the fragments observed in the acquired MS/MS spectrum. For library C, the retention times and elution profiles for all isobaric peptide sequences reported for an MS/MS event were additionally compared with the retention times and elution profiles of all of the isobaric peptides in the sample of the original peptide library. For that reason,

retention times for the analysis of the original and selected library C were aligned using the software POSTMan.¹⁷ By this manual data inspection, several top ranking hits were excluded as false positive hits whereas other hits were included which originally were not assigned by top scores in the Mascot search.

4.9. Analysis of amino acid preference at given positions

Relative distribution (RD) values, given as the frequency of an amino acid at a dedicated Z position multiplied by the total number of amino acids represented at the same position, were calculated to visualize amino acid preference. At a given position, favored amino acids have RD values larger than one, and non-favored amino acids have RD values smaller than 1.

References and notes

1. Sollid, L. M.; Markussen, G.; Ek, J.; Gjerde, H.; Vartdal, F.; Thorsby, E. *J.Exp.Med.* **1989**, *169*, 345-350.
2. Todd, J. A.; Bell, J. I.; McDevitt, H. O. *Nature* **1987**, *329*, 599-604.
3. Thorsby, E. *Hum.Immunol.* **1997**, *53*, 1-11.
4. Lundin, K. E.; Scott, H.; Hansen, T.; Paulsen, G.; Halstensen, T. S.; Fausa, O.; Thorsby, E.; Sollid, L. M. *J.Exp.Med.* **1993**, *178*, 187-196.
5. Sollid, L. M. *Nat.Rev.Immunol.* **2002**, *2*, 647-655.
6. Sollid, L. M.; Thorsby, E. *Gastroenterology* **1993**, *105*, 910-922.
7. Xia, J.; Bergseng, E.; Fleckenstein, B.; Siegel, M.; Kim, C. Y.; Khosla, C.; Sollid, L. M. *Bioorg.Med.Chem.* **2007**, *15*, 6565-6573.
8. Kapoerchan, V. V.; Wiesner, M.; Overhand, M.; van der Marel, G. A.; Koning, F.; Overkleef, H. S. *Bioorg.Med.Chem.* **2008**, *16*, 2053-2062.
9. Thorsby, E.; Lie, B. A. *Transpl.Immunol.* **2005**, *14*, 175-182.
10. Vartdal, F.; Johansen, B. H.; Friede, T.; Thorpe, C. J.; Stevanovic, S.; Eriksen, J. E.; Sletten, K.; Thorsby, E.; Rammensee, H. G.; Sollid, L. M. *Eur.J.Immunol.* **1996**, *26*, 2764-2772.
11. van de Wal, Y.; Kooy, Y. M.; Drijfhout, J. W.; Amons, R.; Papadopoulos, G. K.; Koning, F. *Immunogenetics* **1997**, *46*, 484-492.
12. Arentz-Hansen, H.; Körner, R.; Molberg, Ø.; Quarsten, H.; Vader, W.; Kooy, Y. M.; Lundin, K. E.; Koning, F.; Roepstorff, P.; Sollid, L. M.; McAdam, S. N. *J.Exp.Med.* **2000**, *191*, 603-612.
13. Arentz-Hansen, H.; McAdam, S. N.; Molberg, Ø.; Fleckenstein, B.; Lundin, K. E.; Jørgensen, T. J.; Jung, G.; Roepstorff, P.; Sollid, L. M. *Gastroenterology* **2002**, *123*, 803-809.
14. Sjöström, H.; Lundin, K. E.; Molberg, Ø.; Körner, R.; McAdam, S. N.; Anthonsen, D.; Quarsten, H.; Noren, O.; Roepstorff, P.; Thorsby, E.; Sollid, L. M. *Scand.J.Immunol.* **1998**, *48*, 111-115.
15. Jüse, U.; van de, W. Y.; Koning, F.; Sollid, L. M.; Fleckenstein, B. *Hum.Immunol.* **2010**, *Epub ahead of print*,
<http://dx.doi.org/10.1016/j.humimm.2010.01.021>.
16. Quarsten, H.; McAdam, S. N.; Jensen, T.; Arentz-Hansen, H.; Molberg, Ø.; Lundin, K. E.; Sollid, L. M. *J.Immunol.* **2001**, *167*, 4861-4868.
17. Arntzen, M. O.; Osland, C. L.; Raa, C. R.; Kopperud, R.; Doskeland, S. O.; Lewis, A. E.; D'Santos, C. S. *Proteomics*. **2009**, *9*, 1400-1406.

Table 1

Peptide libraries used in this study. Libraries A and B carry each a single and fully randomized X position which is composed of an equimolar mixture of 19 proteinogenic amino acids (excluding cysteine). Libraries C and D represent dedicated peptide libraries with four and six dedicated Z positions whose composition are given below.

| peptide binding register | | | | | | | | | | | | [sequences] |
|--------------------------|----|----------|---|---|----------|---|----------|----------|----------|----------|-------------------|-------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | |
| Library A | GG | F | A | P | E | K | E | E | P | X | R-NH ₂ | 19 |
| Library B | GG | X | A | P | E | K | E | E | P | F | R-NH2 | 19 |
| Library C | Ac | A | A | A | A | K | Z | Z | Z | Z | R-NH2 | 1430 |
| Library D | Ac | Z | A | A | Z | K | Z | Z | Z | Z | R-NH2 | 64 350 |

Amino acid composition in the Z positions of dedicated libraries C and D:

Z in position 1: P, V, E, W, F

Z in position 4: P, V, T, G, W, L, Y, N, E

Z in position 6: Q, T, N, Y, F, V, G, L, A, P, E

Z in position 7: G, L, F, A, E

Z in position 8: P, E

Z in position 9: G, Q, A, S, V, P, L, E, Y, D, M, F, W

Table 2

Number of MS/MS events and identified peptides in experiments with libraries C and D.

| | Library C | Library D |
|--|-----------|-----------|
| Theoretical sequences | 1430 | 64 350 |
| MS/MS events acquired before incubation | 5059 | 5791 |
| MS/MS events acquired in the eluted sample | 707 | 1212 |
| Sequences identified before incubation | 669 | 1985 |
| Sequences identified in the eluted sample | 81 | 261 |

Figure legends

Fig. 1

Schematic representation of peptide library D registering into the peptide binding groove of HLA-DQ2.5. Five of the six dedicated Z positions interact with the five pockets in the groove and one is bridging a ridge in position P8. The positively charged amino acid lysine is placed in the non-anchor position P5.

Fig. 2

MALDI-TOF mass spectrum of the original library A (top) and of the sample obtained after incubation of HLA-DQ2.5 with library A and peptide elution (bottom). Signals are assigned by the amino acid in the randomized position P9 of the identified peptide sequence.

Fig. 3

Relative distribution (RD) of amino acids present in the dedicated positions of peptide library C before incubation (A) and after elution (B).

Fig. 4

LC-MS map of library C before incubation with HLA-DQ2.5 (left panels) and after elution from HLA-DQ2.5 (right panels). The upper panels show the elution profiles in the region m/z 425-725 and with retention times from 41 min to 73 min. The lower panels present an enlarged 3D view of the rectangular boxes included in the upper panels. Before incubation with HLA-DQ2.5, the lower panel shows three peptides with slightly different retention times (observed as doubly charged ions). The peptides

eluting at 50.1 min and 51.6 min are isobaric with m/z 589.9 and MS/MS data allowed identifying their sequences as AAAAKEEEYR and AAAAKYEEER, respectively. After incubation with HLA-DQ2.5 only one peptide was observed at an aligned retention time of 50.4 min which was identified as AAAAKEEEYR. Images were created with MSight version 2.A.6 and aligned retention times were calculated using POSTMan.

Fig. 5

The range of $\sum relC$ values for all peptides present in library C and library D. The average of $\sum relC$ values of the original library (***) and of the peptides eluted from HLA-DQ2.5 (*) are given.

Fig. 6

Relative distribution (RD) of amino acids present in the dedicated positions of peptide library D before incubation (A) and after elution (B). In order to compensate for a potential bias in these data, RD values given in B were divided by RD values given in A resulting in adjusted RD values (C).

Fig. 7

The amino acids in the dedicated positions P1, P4, P6, P7, P8, and P9 in library C and D were chosen based on their relative competition values ($relC$).¹⁵ Unfavorable amino acids have low $relC$ values; favorable amino acids have high $relC$ values. Residues included in dedicated positions of library C and D are indicated in gray. In each dedicated position, mainly intermediate and good residues in the corresponding position with a $relC$ value greater 1.5 were included (except L in position P7 and P in

position 8). The most abundant residues identified in the peptides eluted from HLA-DQ2.5 after incubation are marked with stars (library C = *, library D = **). By comparing RD values of the eluted sample of library D with the RD values of the sample before incubation, some additional amino acids are more abundant in the eluted sample than before incubation. These amino acids are labeled with (**). In library C, tryptophan (W) in position P9 is less abundant in the eluted sample than in the sample before incubation (*).

Fig. 1

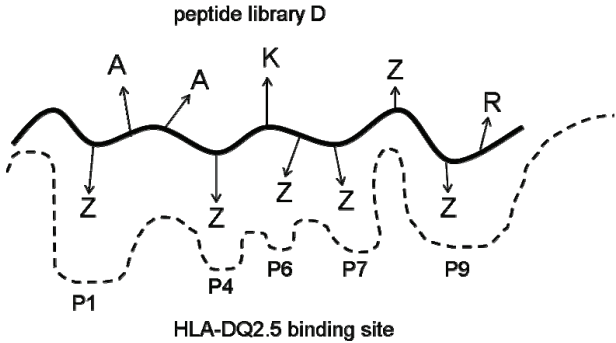


Fig. 2

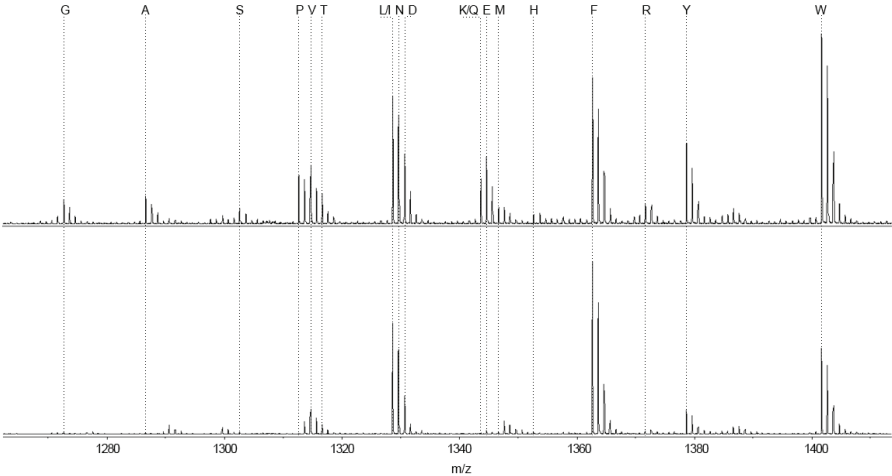


Fig. 3

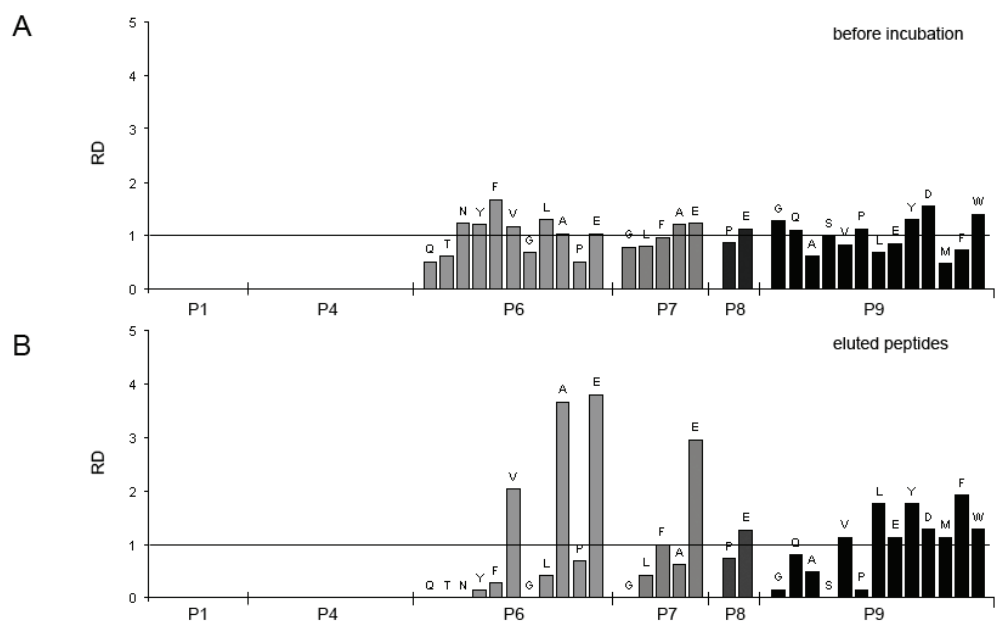


Fig. 4

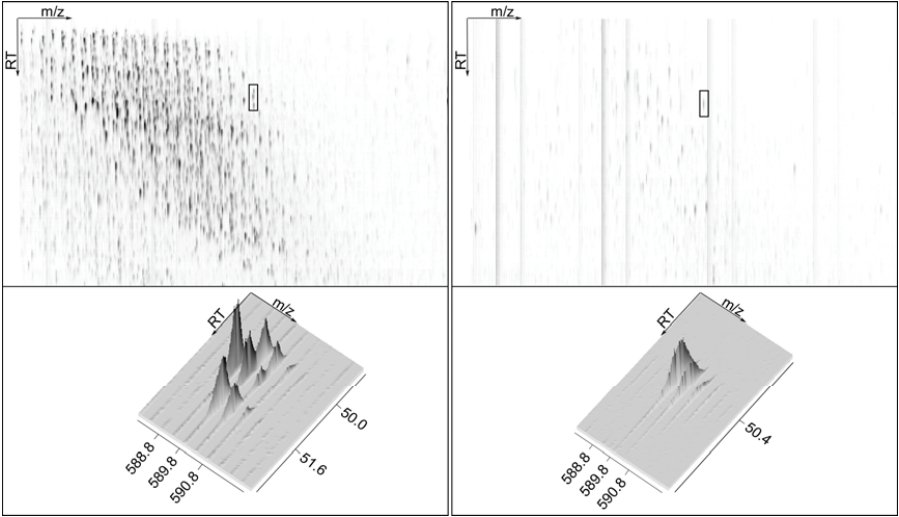


Fig. 5

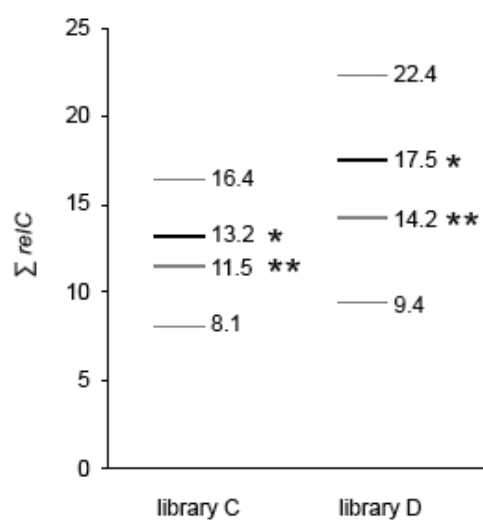


Fig. 6

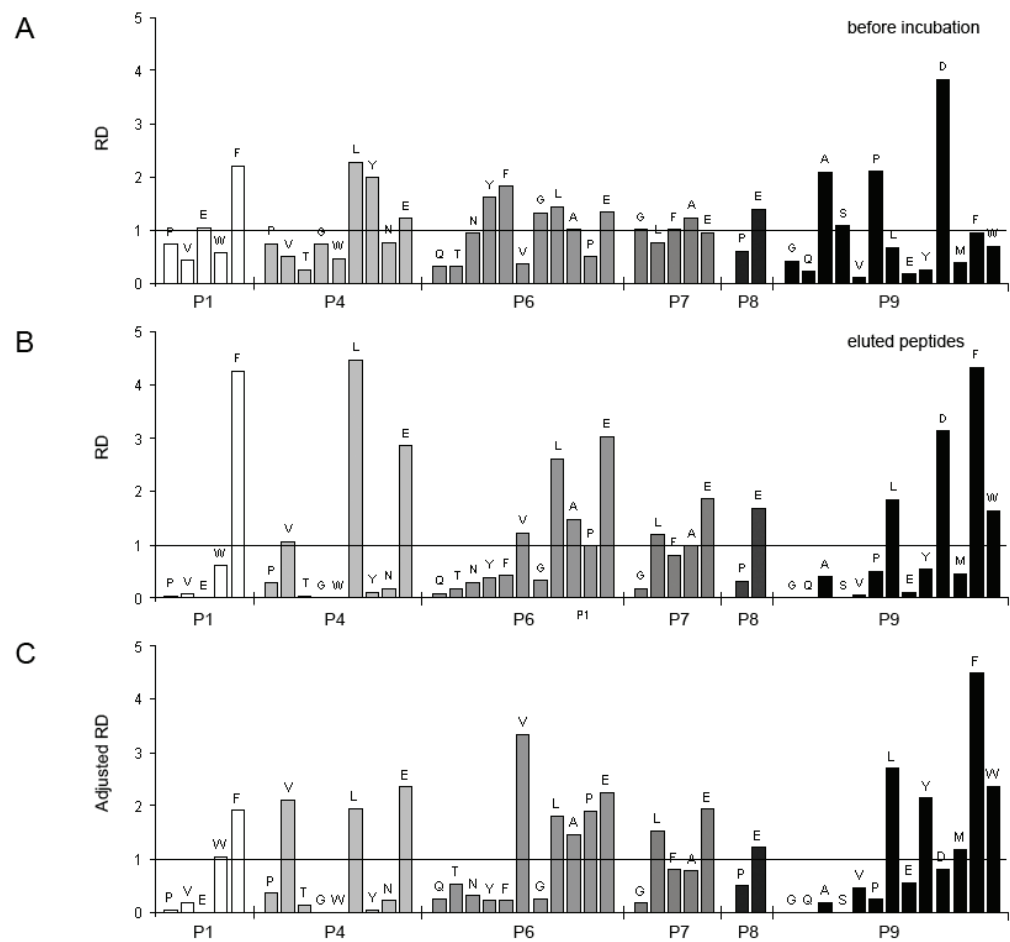


Fig. 7

